(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 December 2001 (27.12.2001)

PCT

(10) International Publication Number
WO 01/98468 A2

(51) International Patent Classification[7]: C12N 9/00

(21) International Application Number: PCT/US01/19178

(22) International Filing Date: 13 June 2001 (13.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/212,336	16 June 2000 (16.06.2000)	US
60/213,955	22 June 2000 (22.06.2000)	US
60/215,396	29 June 2000 (29.06.2000)	US
60/216,821	7 July 2000 (07.07.2000)	US
60/218,946	14 July 2000 (14.07.2000)	US

560 20 N0:4

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). **LAL, Preeti** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **AU-YOUNG, Janice** [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). **DELEGEANE, Angelo, M.** [US/US]; 594 Angus Drive, Milpitas, CA 95035 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **NGUYEN, Danniel, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **LEE, Ernestine, A.** [US/US]; 624 Kains Street, Albany, CA 94706 (US). **HAFALIA, April** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **KHAN, Farrah, A.** [IN/US]; 3617 Central Road #102, Glenview, IL 60025 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). **YAO, Monique, G.** [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **PATTERSON, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **WALSH, Roderick, T.** [IE/US]; 8 Boundary Court, St. Lawrence Road, Canterbury, Kent CT1 3EZ (GB). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **XU, Yuming** [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). **REDDY, Roopa** [IN/US]; 1233 W. McKinley Avenue #3, Sunnyvale, CA 94086 (US). **DAS, Debopriya** [IN/US]; 1179 Bonita Avenue, Apt. 3, Mountain View, CA 94040 (US). **KEARNEY, Liam** [IE/US]; 50 Woodside Avenue, San Francisco, CA 94127 (US). **KALLICK, Deborah, A.** [US/US]; 900 Olive Street, Menlo Park, CA 94025 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: PROTEASES

(57) **Abstract:** The invention provides human proteases (PRTS) and polynucleotides which identify and encode PRTS. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PRTS.

WO 01/98468 A2

PROTEASES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of proteases and to the use of these sequences in hydrolysis of peptide bonds and in the diagnosis, treatment, and prevention of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of proteases.

BACKGROUND OF THE INVENTION

Proteases cleave proteins and peptides at the peptide bond that forms the backbone of the protein or peptide chain. Proteolysis is one of the most important and frequent enzymatic reactions that occurs both within and outside of cells. Proteolysis is responsible for the activation and maturation of nascent polypeptides, the degradation of misfolded and damaged proteins, and the controlled turnover of peptides within the cell. Proteases participate in digestion, endocrine function, and tissue remodeling during embryonic development, wound healing, and normal growth. Proteases can play a role in regulatory processes by affecting the half life of regulatory proteins. Proteases are involved in the etiology or progression of disease states such as inflammation, angiogenesis, tumor dispersion and metastasis, cardiovascular disease, neurological disease, and bacterial, parasitic, and viral infections.

Proteases can be categorized on the basis of where they cleave their substrates. Exopeptidases, which include aminopeptidases, dipeptidyl peptidases, tripeptidases, carboxypeptidases, peptidyl-di-peptidases, dipeptidases, and omega peptidases, cleave residues at the termini of their substrates. Endopeptidases, including serine proteases, cysteine proteases, and metalloproteases, cleave at residues within the peptide. Four principal categories of mammalian proteases have been identified based on active site structure, mechanism of action, and overall three-dimensional structure. (See Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp. 1-5.)

Serine Proteases

The serine proteases (SPs) are a large, widespread family of proteolytic enzymes that include the digestive enzymes trypsin and chymotrypsin, components of the complement and blood-clotting cascades, and enzymes that control the degradation and turnover of macromolecules within the cell and in the extracellular matrix. Most of the more than 20 subfamilies can be grouped into six clans, each with a common ancestor. These six clans are hypothesized to have descended from at least four evolutionarily distinct ancestors. SPs are named for the presence of a serine residue found in the active

catalytic site of most families. The active site is defined by the catalytic triad, a set of conserved asparagine, histidine, and serine residues critical for catalysis. These residues form a charge relay network that facilitates substrate binding. Other residues outside the active site form an oxyanion hole that stabilizes the tetrahedral transition intermediate formed during catalysis. SPs have a wide range of substrates and can be subdivided into subfamilies on the basis of their substrate specificity. The main subfamilies are named for the residue(s) after which they cleave: trypases (after arginine or lysine), aspases (after aspartate), chymases (after phenylalanine or leucine), metases (methionine), and serases (after serine) (Rawlings, N.D. and A.J. Barrett (1994) Methods Enzymol. 244:19-61).

Most mammalian serine proteases are synthesized as zymogens, inactive precursors that are activated by proteolysis. For example, trypsinogen is converted to its active form, trypsin, by enteropeptidase. Enteropeptidase is an intestinal protease that removes an N-terminal fragment from trypsinogen. The remaining active fragment is trypsin, which in turn activates the precursors of the other pancreatic enzymes. Likewise, proteolysis of prothrombin, the precursor of thrombin, generates three separate polypeptide fragments. The N-terminal fragment is released while the other two fragments, which comprise active thrombin, remain associated through disulfide bonds.

The two largest SP subfamilies are the chymotrypsin (S1) and subtilisin (S8) families. Some members of the chymotrypsin family contain two structural domains unique to this family. Kringle domains are triple-looped, disulfide cross-linked domains found in varying copy number. Kringles are thought to play a role in binding mediators such as membranes, other proteins or phospholipids, and in the regulation of proteolytic activity (PROSITE PDOC00020). Apple domains are 90 amino-acid repeated domains, each containing six conserved cysteines. Three disulfide bonds link the first and sixth, second and fifth, and third and fourth cysteines (PROSITE PDOC00376). Apple domains are involved in protein-protein interactions. S1 family members include trypsin, chymotrypsin, coagulation factors IX-XII, complement factors B, C, and D, granzymes, kallikrein, and tissue- and urokinase-plasminogen activators. The subtilisin family has members found in the eubacteria, archaebacteria, eukaryotes, and viruses. Subtilisins include the proprotein-processing endopeptidases kexin and furin and the pituitary prohormone convertases PC1, PC2, PC3, PC6, and PACE4 (Rawlings and Barrett, supra). The prolyl oligopeptidase (S9) family includes enzymes from prokaryotes and eukaryotes with greatly differing specificities. Dipeptidyl peptidase IV (DPP-IV) is identical to CD26 and is implicated in the inactivation of peptide hormones, as well as in regulating T-cell growth (reviewed in Kahne, T. et al. (1999) Int. J. Mol. Med. 4:3-15; Mentlein, R. (1999) Regul. Pept. 85:9-24). Inhibition of DPP-IV has been suggested as a treatment for type 2 diabetes (Holst, J.J. and C.F. Deacon (1998) Diabetes 47:1663-1670), and lowered serum DPP-IV activity has been measured in anorexia and bulimia patients (van West, D. et al. (2000) Eur. Arch. Psych. Clin. Neurosci. 250:86-92).

SPs have functions in many normal processes and some have been implicated in the etiology or treatment of disease. Enterokinase, the initiator of intestinal digestion, is found in the intestinal brush border, where it cleaves the acidic propeptide from trypsinogen to yield active trypsin (Kitamoto, Y. et al. (1994) Proc. Natl. Acad. Sci. USA 91:7588-7592). Prolylcarboxypeptidase, a lysosomal serine peptidase that cleaves peptides such as angiotensin II and III and [des-Arg9] bradykinin, shares sequence homology with members of both the serine carboxypeptidase and prolylendopeptidase families (Tan, F. et al. (1993) J. Biol. Chem. 268:16631-16638). The protease neuropsin may influence synapse formation and neuronal connectivity in the hippocampus in response to neural signaling (Chen, Z.-L. et al. (1995) J. Neurosci. 15:5088-5097). Tissue plasminogen activator is useful for acute management of stroke (Zivin, J.A. (1999) Neurology 53:14-19) and myocardial infarction (Ross, A.M. (1999) Clin. Cardiol. 22:165-171). Some receptors (PAR, for proteinase-activated receptor), highly expressed throughout the digestive tract, are activated by proteolytic cleavage of an extracellular domain. The major agonists for PARs, thrombin, trypsin, and mast cell tryptase, are released in allergy and inflammatory conditions. Control of PAR activation by proteases has been suggested as a promising therapeutic target (Vergnolle, N. (2000) Aliment. Pharmacol. Ther. 14:257-266; Rice, K.D. et al. (1998) Curr. Pharm. Des. 4:381-396). Prostate-specific antigen (PSA) is a kallikrein-like serine protease synthesized and secreted exclusively by epithelial cells in the prostate gland. Serum PSA is elevated in prostate cancer and is the most sensitive physiological marker for monitoring cancer progression and response to therapy. PSA can also identify the prostate as the origin of a metastatic tumor (Brawer, M.K. and P.H. Lange (1989) Urology 33:11-16).

The signal peptidase is a specialized class of SP found in all prokaryotic and eukaryotic cell types that serves in the processing of signal peptides from certain proteins. Signal peptides are amino-terminal domains of a protein which direct the protein from its ribosomal assembly site to a particular cellular or extracellular location. Once the protein has been exported, removal of the signal sequence by a signal peptidase and posttranslational processing, e.g., glycosylation or phosphorylation, activate the protein. Signal peptidases exist as multi-subunit complexes in both yeast and mammals. The canine signal peptidase complex is composed of five subunits, all associated with the microsomal membrane and containing hydrophobic regions that span the membrane one or more times (Shelness, G.S. and G. Blobel (1990) J. Biol. Chem. 265:9512-9519). Some of these subunits serve to fix the complex in its proper position on the membrane while others contain the actual catalytic activity.

Another family of proteases which have a serine in their active site are dependent on the hydrolysis of ATP for their activity. These proteases contain proteolytic core domains and regulatory ATPase domains which can be identified by the presence of the P-loop, an ATP/GTP-binding motif (PROSITE PDOC00803). Members of this family include the eukaryotic mitochondrial matrix

proteases, Clp protease and the proteasome. Clp protease was originally found in plant chloroplasts but is believed to be widespread in both prokaryotic and eukaryotic cells. The gene for early-onset torsion dystonia encodes a protein related to Clp protease (Ozelius, L.J. et al. (1998) Adv. Neurol. 78:93-105).

The proteasome is an intracellular protease complex found in some bacteria and in all eukaryotic cells, and plays an important role in cellular physiology. Proteasomes are associated with the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins of all types, including proteins that function to activate or repress cellular processes such as transcription and cell cycle progression (Ciechanover, A. (1994) Cell 79:13-21). In the UCS pathway, proteins targeted for degradation are conjugated to ubiquitin, a small heat stable protein. The ubiquitinated protein is then recognized and degraded by the proteasome. The resultant ubiquitin-peptide complex is hydrolyzed by a ubiquitin carboxyl terminal hydrolase, and free ubiquitin is released for reutilization by the UCS. Ubiquitin-proteasome systems are implicated in the degradation of mitotic cyclic kinases, oncoproteins, tumor suppressor genes (p53), cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, supra). This pathway has been implicated in a number of diseases, including cystic fibrosis, Angelman's syndrome, and Liddle syndrome (reviewed in Schwartz, A.L. and A. Ciechanover (1999) Annu. Rev. Med. 50:57-74). A murine proto-oncogene, Unp, encodes a nuclear ubiquitin protease whose overexpression leads to oncogenic transformation of NIH3T3 cells. The human homologue of this gene is consistently elevated in small cell tumors and adenocarcinomas of the lung (Gray, D.A. (1995) Oncogene 10:2179-2183). Ubiquitin carboxyl terminal hydrolase is involved in the differentiation of a lymphoblastic leukemia cell line to a non-dividing mature state (Maki, A. et al. (1996) Differentiation 60:59-66). In neurons, ubiquitin carboxyl terminal hydrolase (PGP 9.5) expression is strong in the abnormal structures that occur in human neurodegenerative diseases (Lowe, J. et al. (1990) J. Pathol. 161:153-160). The proteasome is a large (~2000 kDa) multisubunit complex composed of a central catalytic core containing a variety of proteases arranged in four seven-membered rings with the active sites facing inwards into the central cavity, and terminal ATPase subunits covering the outer port of the cavity and regulating substrate entry (for review, see Schmidt, M. et al. (1999) Curr. Opin. Chem. Biol. 3:584-591).

Cysteine Proteases

Cysteine proteases (CPs) are involved in diverse cellular processes ranging from the processing of precursor proteins to intracellular degradation. Nearly half of the CPs known are present only in viruses. CPs have a cysteine as the major catalytic residue at the active site where catalysis proceeds via a thioester intermediate and is facilitated by nearby histidine and asparagine residues. A glutamine residue is also important, as it helps to form an oxyanion hole. Two important CP families include the

papain-like enzymes (C1) and the calpains (C2). Papain-like family members are generally lysosomal or secreted and therefore are synthesized with signal peptides as well as propeptides. Most members bear a conserved motif in the propeptide that may have structural significance (Karrer, K.M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:3063-3067). Three-dimensional structures of papain family members show a bilobed molecule with the catalytic site located between the two lobes. Papains include cathepsins B, C, H, L, and S, certain plant allergens and dipeptidyl peptidase (for a review, see Rawlings, N.D. and A.J. Barrett (1994) Methods Enzymol. 244:461-486).

Some CPs are expressed ubiquitously, while others are produced only by cells of the immune system. Of particular note, CPs are produced by monocytes, macrophages and other cells which migrate to sites of inflammation and secrete molecules involved in tissue repair. Overabundance of these repair molecules plays a role in certain disorders. In autoimmune diseases such as rheumatoid arthritis, secretion of the cysteine peptidase cathepsin C degrades collagen, laminin, elastin and other structural proteins found in the extracellular matrix of bones. Bone weakened by such degradation is also more susceptible to tumor invasion and metastasis. Cathepsin L expression may also contribute to the influx of mononuclear cells which exacerbates the destruction of the rheumatoid synovium (Keyszer, G.M. (1995) Arthritis Rheum. 38:976-984).

Calpains are calcium-dependent cytosolic endopeptidases which contain both an N-terminal catalytic domain and a C-terminal calcium-binding domain. Calpain is expressed as a proenzyme heterodimer consisting of a catalytic subunit unique to each isoform and a regulatory subunit common to different isoforms. Each subunit bears a calcium-binding EF-hand domain. The regulatory subunit also contains a hydrophobic glycine-rich domain that allows the enzyme to associate with cell membranes. Calpains are activated by increased intracellular calcium concentration, which induces a change in conformation and limited autolysis. The resultant active molecule requires a lower calcium concentration for its activity (Chan, S.L. and M.P. Mattson (1999) J. Neurosci. Res. 58:167-190). Calpain expression is predominantly neuronal, although it is present in other tissues. Several chronic neurodegenerative disorders, including ALS, Parkinson's disease and Alzheimer's disease are associated with increased calpain expression (Chan and Mattson, supra). Calpain-mediated breakdown of the cytoskeleton has been proposed to contribute to brain damage resulting from head injury (McCracken, E. et al. (1999) J. Neurotrauma 16:749-761). Calpain-3 is predominantly expressed in skeletal muscle, and is responsible for limb-girdle muscular dystrophy type 2A (Minami, N. et al. (1999) J. Neurol. Sci. 171:31-37).

Another family of thiol proteases is the caspases, which are involved in the initiation and execution phases of apoptosis. A pro-apoptotic signal can activate initiator caspases that trigger a proteolytic caspase cascade, leading to the hydrolysis of target proteins and the classic apoptotic death

of the cell. Two active site residues, a cysteine and a histidine, have been implicated in the catalytic mechanism. Caspases are among the most specific endopeptidases, cleaving after aspartate residues. Caspases are synthesized as inactive zymogens consisting of one large (p20) and one small (p10) subunit separated by a small spacer region, and a variable N-terminal prodomain. This prodomain interacts with cofactors that can positively or negatively affect apoptosis. An activating signal causes autoproteolytic cleavage of a specific aspartate residue (D297 in the caspase-1 numbering convention) and removal of the spacer and prodomain, leaving a p10/p20 heterodimer. Two of these heterodimers interact via their small subunits to form the catalytically active tetramer. The long prodomains of some caspase family members have been shown to promote dimerization and auto-processing of procaspases. Some caspases contain a "death effector domain" in their prodomain by which they can be recruited into self-activating complexes with other caspases and FADD protein associated death receptors or the TNF receptor complex. In addition, two dimers from different caspase family members can associate, changing the substrate specificity of the resultant tetramer. Endogenous caspase inhibitors (inhibitor of apoptosis proteins, or IAPs) also exist. All these interactions have clear effects on the control of apoptosis (reviewed in Chan and Mattson, supra; Salveson, G.S. and V.M. Dixit (1999) Proc. Natl. Acad. Sci. USA 96:10964-10967).

Caspases have been implicated in a number of diseases. Mice lacking some caspases have severe nervous system defects due to failed apoptosis in the neuroepithelium and suffer early lethality. Others show severe defects in the inflammatory response, as caspases are responsible for processing IL-1b and possibly other inflammatory cytokines (Chan and Mattson, supra). Cowpox virus and baculoviruses target caspases to avoid the death of their host cell and promote successful infection. In addition, increases in inappropriate apoptosis have been reported in AIDS, neurodegenerative diseases and ischemic injury, while a decrease in cell death is associated with cancer (Salveson and Dixit, supra; Thompson, C.B. (1995) Science 267:1456-1462).

Aspartyl proteases

Aspartyl proteases (APs) include the lysosomal proteases cathepsins D and E, as well as chymosin, renin, and the gastric pepsins. Most retroviruses encode an AP, usually as part of the pol polyprotein. APs, also called acid proteases, are monomeric enzymes consisting of two domains, each domain containing one half of the active site with its own catalytic aspartic acid residue. APs are most active in the range of pH 2–3, at which one of the aspartate residues is ionized and the other neutral. The pepsin family of APs contains many secreted enzymes, and all are likely to be synthesized with signal peptides and propeptides. Most family members have three disulfide loops, the first ~5 residue loop following the first aspartate, the second 5-6 residue loop preceding the second aspartate, and the third and largest loop occurring toward the C terminus. Retropepsins, on the other hand, are analogous

to a single domain of pepsin, and become active as homodimers with each retropepsin monomer contributing one half of the active site. Retropepsins are required for processing the viral polyproteins.

APs have roles in various tissues, and some have been associated with disease. Renin mediates the first step in processing the hormone angiotensin, which is responsible for regulating electrolyte balance and blood pressure (reviewed in Crews, D.E. and S.R. Williams (1999) Hum. Biol. 71:475-503). Abnormal regulation and expression of cathepsins are evident in various inflammatory disease states. Expression of cathepsin D is elevated in synovial tissues from patients with rheumatoid arthritis and osteoarthritis. The increased expression and differential regulation of the cathepsins are linked to the metastatic potential of a variety of cancers (Chambers, A.F. et al. (1993) Crit. Rev. Oncol. 4:95-114).

Metalloproteases

Metalloproteases require a metal ion for activity, usually manganese or zinc. Examples of manganese metalloenzymes include aminopeptidase P and human proline dipeptidase (PEPD). Aminopeptidase P can degrade bradykinin, a nonapeptide activated in a variety of inflammatory responses. Aminopeptidase P has been implicated in coronary ischemia/reperfusion injury. Administration of aminopeptidase P inhibitors has been shown to have a cardioprotective effect in rats (Ersahin, C. et al (1999) J. Cardiovasc. Pharmacol. 34:604-611).

Most zinc-dependent metalloproteases share a common sequence in the zinc-binding domain. The active site is made up of two histidines which act as zinc ligands and a catalytic glutamic acid C-terminal to the first histidine. Proteins containing this signature sequence are known as the metzincins and include aminopeptidases B and N, angiotensin-converting enzyme, neurolysin, the matrix metalloproteases and the adamalysins (ADAMS). An alternate sequence is found in the zinc carboxypeptidases, in which all three conserved residues – two histidines and a glutamic acid – are involved in zinc binding.

A number of the neutral metalloendopeptidases, including angiotensin converting enzyme and the aminopeptidases, are involved in the metabolism of peptide hormones. High aminopeptidase B activity, for example, is found in the adrenal glands and neurohypophyses of hypertensive rats (Prieto, I. et al. (1998) Horm. Metab. Res. 30:246-248). Oligopeptidase M/neurolysin can hydrolyze bradykinin as well as neurotensin (Serizawa, A. et al. (1995) J. Biol. Chem 270:2092-2098). Neurotensin is a vasoactive peptide that can act as a neurotransmitter in the brain, where it has been implicated in limiting food intake (Tritos, N.A. et al. (1999) Neuropeptides 33:339-349).

The matrix metalloproteases (MMPs) are a family of at least 23 enzymes that can degrade components of the extracellular matrix (ECM). They are Zn^{+2} endopeptidases with an N-terminal catalytic domain. Nearly all members of the family have a hinge peptide and C-terminal domain which

can bind to substrate molecules in the ECM or to inhibitors produced by the tissue (TIMPs, for tissue inhibitor of metalloprotease; Campbell, I.L. et al. (1999) Trends Neurosci. 22:285). The presence of fibronectin-like repeats, transmembrane domains, or C-terminal hemopexinase-like domains can be used to separate MMPs into collagenase, gelatinase, stromelysin and membrane-type MMP subfamilies. In the inactive form, the Zn^{+2} ion in the active site interacts with a cysteine in the pro-sequence. Activating factors disrupt the Zn^{+2}-cysteine interaction, or "cysteine switch," exposing the active site. This partially activates the enzyme, which then cleaves off its propeptide and becomes fully active. MMPs are often activated by the serine proteases plasmin and furin. MMPs are often regulated by stoichiometric, noncovalent interactions with inhibitors; the balance of protease to inhibitor, then, is very important in tissue homeostasis (reviewed in Yong, V.W. et al. (1998) Trends Neurosci. 21:75). Ehlers-Danlos syndrome type VII C is caused by mutations in the procollagen I N-proteinase gene (Colige, A. et al. (1999) Am. J. Hum. Genet. 65:308-317).

MMPs are implicated in a number of diseases including osteoarthritis (Mitchell, P. et al. (1996) J. Clin. Invest. 97:761), atherosclerotic plaque rupture (Sukhova, G.K. et al. (1999) Circulation 99:2503), aortic aneurysm (Schneiderman, J. et al. (1998) Am. J. Path. 152:703), non-healing wounds (Saarialho-Kere, U.K. et al. (1994) J. Clin. Invest. 94:79), bone resorption (Blavier, L. and J.M. Delaisse (1995) J. Cell Sci. 108:3649), age-related macular degeneration (Steen, B. et al. (1998) Invest. Ophthalmol. Vis. Sci. 39:2194), emphysema (Finlay, G.A. et al. (1997) Thorax 52:502), myocardial infarction (Rohde, L.E. et al. (1999) Circulation 99:3063) and dilated cardiomyopathy (Thomas, C.V. et al. (1998) Circulation 97:1708). MMP inhibitors prevent metastasis of mammary carcinoma and experimental tumors in rat, and Lewis lung carcinoma, hemangioma, and human ovarian carcinoma xenografts in mice (Eccles, S.A. et al. (1996) Cancer Res. 56:2815; Anderson et al. (1996) Cancer Res. 56:715-718; Volpert, O.V. et al. (1996) J. Clin. Invest. 98:671; Taraboletti, G. et al. (1995) J. NCI 87:293; Davies, B. et al. (1993) Cancer Res. 53:2087). MMPs may be active in Alzheimer's disease. A number of MMPs are implicated in multiple sclerosis, and administration of MMP inhibitors can relieve some of its symptoms (reviewed in Yong, supra).

Another family of metalloproteases is the ADAMs, for A Disintegrin and Metalloprotease Domain, which they share with their close relatives the adamalysins, snake venom metalloproteases (SVMPs). ADAMs combine features of both cell surface adhesion molecules and proteases, containing a prodomain, a protease domain, a disintegrin domain, a cysteine rich domain, an epidermal growth factor repeat, a transmembrane domain, and a cytoplasmic tail. The first three domains listed above are also found in the SVMPs. The ADAMs possess four potential functions: proteolysis, adhesion, signaling and fusion. The ADAMs share the metzincin zinc binding sequence and are inhibited by some MMP antagonists such as TIMP-1.

ADAMs are implicated in such processes as sperm-egg binding and fusion, myoblast fusion, and protein-ectodomain processing or shedding of cytokines, cytokine receptors, adhesion proteins and other extracellular protein domains (Schlöndorff, J. and C.P. Blobel (1999) J. Cell. Sci. 112:3603-3617). The Kuzbanian protein cleaves a substrate in the NOTCH pathway (possibly NOTCH itself), activating the program for lateral inhibition in Drosophila neural development. Two ADAMs, TACE (ADAM 17) and ADAM 10, are proposed to have analogous roles in the processing of amyloid precursor protein in the brain (Schlöndorff and Blobel, supra). TACE has also been identified as the TNF activating enzyme (Black, R.A. et al. (1997) Nature 385:729). TNF is a pleiotropic cytokine that is important in mobilizing host defenses in response to infection or trauma, but can cause severe damage in excess and is often overproduced in autoimmune disease. TACE cleaves membrane-bound pro-TNF to release a soluble form. Other ADAMs may be involved in a similar type of processing of other membrane-bound molecules.

The ADAMTS sub-family has all of the features of ADAM family metalloproteases and contain an additional thrombospondin domain (TS). The prototypic ADAMTS was identified in mouse, found to be expressed in heart and kidney and upregulated by proinflammatory stimuli (Kuno, K. et al. (1997) J. Biol. Chem. 272:556-562). To date eleven members are recognized by the Human Genome Organization (HUGO; http://www.gene.ucl.ac.uk/users/hester/adamts.html#Approved). Members of this family have the ability to degrade aggrecan, a high molecular weight proteoglycan which provides cartilage with important mechanical properties including compressibility, and which is lost during the development of arthritis. Enzymes which degrade aggrecan are thus considered attractive targets to prevent and slow the degradation of articular cartilage (See, e.g., Tortorella, M.D. (1999) Science 284:1664; Abbaszade, I. (1999) J. Biol. Chem. 274:23443). Other members are reported to have antiangiogenic potential (Kuno et al., supra) and/or procollagen processing (Colige, A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2374).

The discovery of new proteases and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in hydrolysis of peptide bonds and in the diagnosis, prevention, and treatment of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of proteases.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, proteases, referred to collectively as "PRTS" and individually as "PRTS-1," "PRTS-2," "PRTS-3," "PRTS-4," "PRTS-5," "PRTS-6," "PRTS-7,"

"PRTS-8," "PRTS-9," "PRTS-10," "PRTS-11," "PRTS-12," "PRTS-13," "PRTS-14," "PRTS-15," "PRTS-16," "PRTS-17," "PRTS-18," "PRTS-19," "PRTS-20," and "PRTS-21." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-21.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-21. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:22-42.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group

consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a

polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an

amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PRTS, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:22-42, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"PRTS" refers to the amino acid sequences of substantially purified PRTS obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PRTS. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PRTS either by directly interacting with PRTS or by acting on components of the biological pathway in which PRTS participates.

An "allelic variant" is an alternative form of the gene encoding PRTS. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding PRTS include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PRTS or a polypeptide with at least one functional characteristic of PRTS. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PRTS, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PRTS. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PRTS. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PRTS is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic

molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of PRTS. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PRTS either by directly interacting with PRTS or by acting on components of the biological pathway in which PRTS participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind PRTS polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or

translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PRTS, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PRTS or fragments of PRTS may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu

Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

A "fragment" is a unique portion of PRTS or the polynucleotide encoding PRTS which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10,

15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:22-42 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:22-42, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:22-42 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:22-42 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:22-42 and the region of SEQ ID NO:22-42 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-21 is encoded by a fragment of SEQ ID NO:22-42. A fragment of SEQ ID NO:1-21 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-21. For example, a fragment of SEQ ID NO:1-21 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-21. The precise length of a fragment of SEQ ID NO:1-21 and the region of SEQ ID NO:1-21 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular

biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) <u>Molecular Cloning: A Laboratory Manual</u>, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily

apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PRTS which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PRTS which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of PRTS. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PRTS.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding

sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an PRTS may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PRTS.

"Probe" refers to nucleic acid sequences encoding PRTS, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing PRTS, nucleic acids encoding PRTS, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may

possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human proteases (PRTS), the polynucleotides encoding PRTS, and the use of these compositions for the diagnosis, treatment, or prevention of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank

homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are proteases. For example, SEQ ID NO:1 is a ubiquitin carboxyl terminal hydrolase. SEQ ID NO:1 is 48% identical, from residue M1 to residue G225, to human ubiquitin-specific processing protease (GenBank ID g9971757) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.00e-49, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 contains a ubiquitin carboxyl terminal hydrolase catalytic site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. The score is 53.4 bits and the E-value is 4.9e-12, which indicates the probability of obtaining the observed structural motif by chance. The presence of this motif was corroborated by BLIMPS (probability score=2.6e-4) and MOTIFS analyses. This provides further evidence that SEQ ID NO:1 is a ubiquitin carboxyl-terminal hydrolase. In an alternative example, SEQ ID NO:2 is 45% identical to amino acids 15-235 of human prostasin, a serine protease (GenBank ID g1143194) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.3e-46, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains a trypsin family serine protease active site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. This match has a probability score of 2.7e-58. BLIMPS, MOTIFS, and PROFILESCAN analyses confirm the presence of this domain. (See Table 3.) BLIMPS analysis also reveals a kringle domain, providing further corroborative evidence that SEQ ID NO:2 is a serine protease of the trypsin family. In an alternative example, SEQ ID NO:7 is a dipeptidase which hydrolyses a variety of peptides (Kozak, E. and S. Tate (1982) J. Biol. Chem. 257:6322-6327), and is responsible for the hydrolysis of the beta

lactam rings of antibiotics such as penem and carbapenem (Campbell et al., (1984) J. Biol. Chem. 259:14586-14590). SEQ ID NO:7 shows 48% amino acid sequence identity over 377 amino acids (total length equals 411 amino acids) to human dipeptidase precursor (GenBank ID g219600) as determined by Basic Local Alignment Search Tool (BLAST). The BLAST probability score is 1.1e-88, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Additionally, the protease of the invention demonstrates a renal dipeptidase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. The HMM score for the renal dipeptidase PFAM hit is 412.7. Data from BLIMPS, MOTIFS, BLAST-DOMO, and BLAST-PRODOM analyses provide further corroborative evidence that SEQ ID NO:7 is a renal dipeptidase. The BLIMPS-BLOCKS hit scores for localized regions range from 1040-1537. The BLAST-DOMO hit probability score is 5.2e-85. The BLAST-PRODOM hit probability score is 4.7e-73. In an alternative example, SEQ ID NO:8 is 86% identical to human transmembrane tryptase (GenBank ID g6103629) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.9e-166, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:8 contains a trypsin family protease active site domain with a probability score of 5.3e-89 as determined by searching for matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. BLIMPS, MOTIFS, and PROFILESCAN analyses confirm the presence of this motif. BLIMPS analysis also shows that SEQ ID NO:8 contains a kringle domain and a type I fibronectin domain. HMMER-based analysis reveals the presence of a transmembrane domain (See Table 3.). Taken together, these analyses show that SEQ ID NO:8 is a transmembrane member of the trypsin family of serine proteases. In an alternative example, SEQ ID NO:17 shares 44% local identity with human membrane-type serine protease 1 (MT-SP1, GenBank ID g6002714) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.1e-94, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:17 contains a trypsin family serine protease active site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) HMM-based analysis also reveals a transmembrane domain near the N-terminus of SEQ ID NO:17. A domain found in the low-density lipoprotein receptor and other proteins, including MT-SP1 (PDOC00929) was also identified in this way. The presence of the trypsin active site motif is confirmed by PROFILESCAN, BLIMPS, and MOTIFS analyses. BLIMPS analysis revealed the presence of kringle and type I fibronectin domains. Taken together, these data provide further corroborative evidence that SEQ ID NO:17 is a transmembrane member of the trypsin family of serine proteases. SEQ ID NO:3-6, SEQ ID NO:9-16,

and SEQ ID NO:18-21 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-21 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:22-42 or that distinguish between SEQ ID NO:22-42 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 7246467T8 is the identification number of an Incyte cDNA sequence, and PROSTMY01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71041539V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g5745066) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, GNN.g7208751_000002_002.edit is the identification number of a Genscan-predicted coding sequence, with g7208751 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL1389845_00001 represents a "stitched" sequence in which 1389845 is the identification number of the cluster of sequences to which the algorithm was applied, and 00001 is the number of the prediction generated by the algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. For example, FL2256251_g7708357_000002_g6103629 is the identification

number of a "stretched" sequence, with 2256251 being the Incyte project identification number, g7708357 being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, and g6103629 being the GenBank identification number of the nearest GenBank protein homolog. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses PRTS variants. A preferred PRTS variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PRTS amino acid sequence, and which contains at least one functional or structural characteristic of PRTS.

The invention also encompasses polynucleotides which encode PRTS. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:22-42, which encodes PRTS. The polynucleotide sequences of SEQ ID NO:22-42, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding PRTS. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PRTS. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:22-42 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:22-42. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PRTS.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PRTS, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made

by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PRTS, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PRTS and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PRTS under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PRTS or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PRTS and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PRTS and PRTS derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PRTS or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:22-42 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics,

Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PRTS may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the

emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PRTS may be cloned in recombinant DNA molecules that direct expression of PRTS, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PRTS.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PRTS-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of PRTS, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding PRTS may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, PRTS itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of PRTS, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active PRTS, the nucleotide sequences encoding PRTS or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PRTS. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PRTS. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PRTS and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PRTS and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in

vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PRTS. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PRTS. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PRTS can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding PRTS into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of PRTS are needed, e.g. for the production of antibodies, vectors which direct high level expression of PRTS may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PRTS. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of PRTS. Transcription of sequences encoding PRTS may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PRTS may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PRTS in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of PRTS in cell lines is preferred. For example, sequences encoding PRTS can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a

selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk⁻* and *apr⁻* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PRTS is inserted within a marker gene sequence, transformed cells containing sequences encoding PRTS can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PRTS under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PRTS and that express PRTS may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PRTS using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal

antibodies reactive to two non-interfering epitopes on PRTS is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PRTS include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PRTS, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PRTS may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PRTS may be designed to contain signal sequences which direct secretion of PRTS through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PRTS may be ligated to a heterologous sequence resulting in translation of a fusion

protein in any of the aforementioned host systems. For example, a chimeric PRTS protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PRTS activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PRTS encoding sequence and the heterologous protein sequence, so that PRTS may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PRTS may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

PRTS of the present invention or fragments thereof may be used to screen for compounds that specifically bind to PRTS. At least one and up to a plurality of test compounds may be screened for specific binding to PRTS. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of PRTS, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PRTS binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express PRTS, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing PRTS or cell membrane fractions which contain PRTS are then contacted with

a test compound and binding, stimulation, or inhibition of activity of either PRTS or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with PRTS, either in solution or affixed to a solid support, and detecting the binding of PRTS to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

PRTS of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PRTS. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PRTS activity, wherein PRTS is combined with at least one test compound, and the activity of PRTS in the presence of a test compound is compared with the activity of PRTS in the absence of the test compound. A change in the activity of PRTS in the presence of the test compound is indicative of a compound that modulates the activity of PRTS. Alternatively, a test compound is combined with an in vitro or cell-free system comprising PRTS under conditions suitable for PRTS activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PRTS may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PRTS or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the

resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding PRTS may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding PRTS can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PRTS is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress PRTS, e.g., by secreting PRTS in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

PRTS are useful for hydrolyzing peptide bonds. Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PRTS and proteases. In addition, the expression of PRTS is closely associated with hemic, neurological, reproductive, endocrine, urogenital, diseased, teratocarcinoma, and tumorous tissues,. Therefore, PRTS appears to play a role in gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders. In the treatment of disorders associated with increased PRTS expression or activity, it is desirable to decrease the expression or activity of PRTS. In the treatment of disorders associated with decreased PRTS expression or activity, it is desirable to increase the expression or activity of PRTS.

Therefore, in one embodiment, PRTS or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS. Examples of such disorders include, but are not limited to, a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-

Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha$_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, atherosclerotic plaque rupture, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, degradation of articular cartilage, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal

gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, bone resorption, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, age-related macular degeneration, and sensorineural hearing loss; an epithelial disorder, such as dyshidrotic eczema, allergic contact dermatitis, keratosis pilaris, melasma, vitiligo, actinic keratosis, basal cell carcinoma, squamous cell carcinoma, seborrheic keratosis, folliculitis, herpes simplex, herpes zoster, varicella, candidiasis, dermatophytosis, scabies, insect bites, cherry angioma, keloid, dermatofibroma, acrochordons, urticaria, transient acantholytic dermatosis, xerosis, eczema, atopic dermatitis, contact dermatitis, hand eczema, nummular eczema, lichen simplex chronicus, asteatotic eczema, stasis dermatitis and stasis ulceration, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, impetigo, ecthyma, dermatophytosis, tinea versicolor, warts, acne vulgaris, acne rosacea, pemphigus vulgaris, pemphigus foliaceus, paraneoplastic pemphigus, bullous pemphigoid, herpes gestationis, dermatitis herpetiformis, linear IgA disease, epidermolysis bullosa acquisita, dermatomyositis, lupus erythematosus, scleroderma and morphea, erythroderma, alopecia, figurate skin lesions, telangiectasias, hypopigmentation, hyperpigmentation, vesicles/bullae, exanthems, cutaneous drug reactions, papulonodular skin lesions, chronic non-healing wounds, photosensitivity diseases, epidermolysis bullosa simplex, epidermolytic hyperkeratosis, epidermolytic and nonepidermolytic palmoplantar keratoderma, ichthyosis bullosa of Siemens, ichthyosis exfoliativa, keratosis palmaris et plantaris, keratosis palmoplantaris, palmoplantar keratoderma, keratosis punctata, Meesmann's corneal dystrophy, pachyonychia congenita, white sponge nevus, steatocystoma multiplex, epidermal nevi/epidermolytic hyperkeratosis type, monilethrix, trichothiodystrophy, chronic hepatitis/cryptogenic cirrhosis, and colorectal hyperplasia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including

kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a reproductive disorder, such as infertility, including tubal disease, ovulatory defects, and endometriosis, a disorder of prolactin production, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia.

In another embodiment, a vector capable of expressing PRTS or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified PRTS in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of PRTS may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those listed above.

In a further embodiment, an antagonist of PRTS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PRTS. Examples of such disorders include, but are not limited to, those gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders described above. In one aspect, an antibody which specifically binds PRTS may be used

directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PRTS.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PRTS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PRTS including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PRTS may be produced using methods which are generally known in the art. In particular, purified PRTS may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PRTS. Antibodies to PRTS may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PRTS or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PRTS have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PRTS amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PRTS may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited

to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PRTS-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g.; Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for PRTS may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PRTS and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PRTS epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PRTS. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of PRTS-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined

for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PRTS epitopes, represents the average affinity, or avidity, of the antibodies for PRTS. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular PRTS epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the PRTS-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PRTS, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PRTS-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding PRTS, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PRTS. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PRTS. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Cli. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et

al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PRTS may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PRTS expression or regulation causes disease, the expression of PRTS from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PRTS are treated by constructing mammalian expression vectors encoding PRTS and introducing these vectors by mechanical means into PRTS-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of PRTS include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PRTS may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus

(RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PRTS from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PRTS expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PRTS under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et

al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PRTS to cells which have one or more genetic abnormalities with respect to the expression of PRTS. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PRTS to target cells which have one or more genetic abnormalities with respect to the expression of PRTS. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PRTS to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PRTS to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on

the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for PRTS into the alphavirus genome in place of the capsid-coding region results in the production of a large number of PRTS-coding RNAs and the synthesis of high levels of PRTS in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of PRTS into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PRTS.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of

candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PRTS. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding PRTS. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased PRTS expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PRTS may be therapeutically useful, and in the treatment of disorders associated with decreased PRTS expression or activity, a compound which specifically promotes expression of the polynucleotide encoding PRTS may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a

library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding PRTS is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PRTS are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding PRTS. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PRTS, antibodies to PRTS, and mimetics, agonists, antagonists, or inhibitors of PRTS.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising PRTS or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PRTS or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PRTS or fragments thereof, antibodies of PRTS, and agonists, antagonists or inhibitors of PRTS, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large

therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind PRTS may be used for the diagnosis of disorders characterized by expression of PRTS, or in assays to monitor patients being treated with PRTS or agonists, antagonists, or inhibitors of PRTS. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PRTS include methods which utilize the antibody and a label to detect PRTS in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PRTS, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PRTS expression. Normal or standard values for PRTS expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to PRTS under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PRTS expressed in subject,

control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PRTS may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of PRTS may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PRTS, and to monitor regulation of PRTS levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PRTS or closely related molecules may be used to identify nucleic acid sequences which encode PRTS. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PRTS, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PRTS encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:22-42 or from genomic sequences including promoters, enhancers, and introns of the PRTS gene.

Means for producing specific hybridization probes for DNAs encoding PRTS include the cloning of polynucleotide sequences encoding PRTS or PRTS derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes <u>in vitro</u> by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PRTS may be used for the diagnosis of disorders associated with expression of PRTS. Examples of such disorders include, but are not limited to, a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease,

Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha$_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, atherosclerotic plaque rupture, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, degradation of articular cartilage, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in

particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, bone resorption, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, age-related macular degeneration, and sensorineural hearing loss; an epithelial disorder, such as dyshidrotic eczema, allergic contact dermatitis, keratosis pilaris, melasma, vitiligo, actinic keratosis, basal cell carcinoma, squamous cell carcinoma, seborrheic keratosis, folliculitis, herpes simplex, herpes zoster, varicella, candidiasis, dermatophytosis, scabies, insect bites, cherry angioma, keloid, dermatofibroma, acrochordons, urticaria, transient acantholytic dermatosis, xerosis, eczema, atopic dermatitis, contact dermatitis, hand eczema, nummular eczema, lichen simplex chronicus, asteatotic eczema, stasis dermatitis and stasis ulceration, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, impetigo, ecthyma, dermatophytosis, tinea versicolor, warts, acne vulgaris, acne rosacea, pemphigus vulgaris, pemphigus foliaceus, paraneoplastic pemphigus, bullous pemphigoid, herpes gestationis, dermatitis herpetiformis, linear IgA disease, epidermolysis bullosa acquisita, dermatomyositis, lupus erythematosus, scleroderma and morphea, erythroderma, alopecia, figurate skin lesions, telangiectasias, hypopigmentation, hyperpigmentation, vesicles/bullae, exanthems, cutaneous drug reactions, papulonodular skin lesions, chronic non-healing wounds, photosensitivity diseases, epidermolysis bullosa simplex, epidermolytic hyperkeratosis, epidermolytic and nonepidermolytic palmoplantar keratoderma, ichthyosis bullosa of Siemens, ichthyosis exfoliativa, keratosis palmaris et plantaris, keratosis palmoplantaris, palmoplantar keratoderma, keratosis punctata, Meesmann's corneal dystrophy, pachyonychia congenita, white sponge nevus, steatocystoma multiplex, epidermal nevi/epidermolytic hyperkeratosis type, monilethrix, trichothiodystrophy, chronic hepatitis/cryptogenic cirrhosis, and colorectal hyperplasia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion

diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a reproductive disorder, such as infertility, including tubal disease, ovulatory defects, and endometriosis, a disorder of prolactin production, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia. The polynucleotide sequences encoding PRTS may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PRTS expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PRTS may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PRTS may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PRTS in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PRTS, a normal or standard profile for expression is established. This may be accomplished by combining

body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PRTS, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PRTS may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding PRTS, or a fragment of a polynucleotide complementary to the polynucleotide encoding PRTS, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PRTS may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PRTS are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable

using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of PRTS include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, PRTS, fragments of PRTS, or antibodies specific for PRTS may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a

given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present

invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term *proteome* refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for PRTS to quantify the levels of PRTS expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the

analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PRTS may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic

linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PRTS on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PRTS, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PRTS and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PRTS, or fragments thereof, and washed. Bound PRTS is then detected by methods well known in the art. Purified PRTS can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PRTS specifically compete with a test compound for binding PRTS. In

this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PRTS.

In additional embodiments, the nucleotide sequences which encode PRTS may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/212,336, U.S. Ser. No. 60/213,995, U.S. Ser. No. 60/215,396, U.S. Ser. No. 60/216,821, and U.S. Ser. No. 60/218,946, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic

oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI

PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of

which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:22-42. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative proteases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode proteases, the encoded polypeptides were analyzed by querying against PFAM models for proteases. Potential proteases were also identified by homology to Incyte cDNA sequences that had been annotated as proteases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

<u>"Stitched" Sequences</u>

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped

to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

<u>"Stretched" Sequences</u>

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of PRTS Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:22-42 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched

SEQ ID NO:22-42 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:25 was mapped to chromosome 5 within the interval from 69.60 to 76.50 centiMorgans. SEQ ID NO:28 was mapped to chromosome 16 within the interval from 81.80 to 84.40 centiMorgans.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1), length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated

as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding PRTS are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding PRTS. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of PRTS Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified

using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:22-42 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may

contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

<u>Tissue or Cell Sample Preparation</u>

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by <u>in vitro</u> transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

<u>Microarray Preparation</u>

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are

amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

<u>Hybridization</u>

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

<u>Detection</u>

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the PRTS-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PRTS. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PRTS. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PRTS-encoding transcript.

XII. Expression of PRTS

Expression and purification of PRTS is achieved using bacterial or virus-based expression systems. For expression of PRTS in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PRTS upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PRTS in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PRTS by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PRTS is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PRTS at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified PRTS obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, XVIII, and XIX, where applicable.

XIII. Functional Assays

PRTS function is assessed by expressing the sequences encoding PRTS at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which

contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PRTS on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding PRTS and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PRTS and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of PRTS Specific Antibodies

PRTS substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PRTS amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-PRTS activity by, for example, binding the peptide or PRTS to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring PRTS Using Specific Antibodies

Naturally occurring or recombinant PRTS is substantially purified by immunoaffinity chromatography using antibodies specific for PRTS. An immunoaffinity column is constructed by covalently coupling anti-PRTS antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PRTS are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PRTS (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PRTS binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PRTS is collected.

XVI. Identification of Molecules Which Interact with PRTS

PRTS, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PRTS, washed, and any wells with labeled PRTS complex are assayed. Data obtained using different concentrations of PRTS are used to calculate values for the number, affinity, and association of PRTS with the candidate molecules.

Alternatively, molecules interacting with PRTS are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

PRTS may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of PRTS Activity

Protease activity is measured by the hydrolysis of appropriate synthetic peptide substrates conjugated with various chromogenic molecules in which the degree of hydrolysis is quantified by spectrophotometric (or fluorometric) absorption of the released chromophore (Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp.25-55). Peptide substrates are designed according to the category of protease activity as endopeptidase (serine, cysteine, aspartic proteases, or metalloproteases), aminopeptidase (leucine aminopeptidase), or carboxypeptidase (carboxypeptidases A and B, procollagen C-proteinase). Commonly used chromogens are 2-naphthylamine, 4-nitroaniline, and furylacrylic acid. For example, arginine-β-napthylamide can be used as a substrate for SEQ ID NO:3 (Fukasawa, K.M. et al. (1996) J. Biol. Chem. 271:30731-30735) and 4-phenylazobenzyloxycarbonyl-Pro-Leu-Gly-Pro-D-Arg can be used as a substrate for SEQ ID NO:4. In an alternative example, a substrate for SEQ ID NO:9 would be 7-amino-4-trifluoromethyl coumarin-Phe-Pro-AFC. Assays are performed at ambient temperature and contain an aliquot of the enzyme and the appropriate substrate in a suitable buffer. Reactions are carried out in an optical cuvette, and the increase/decrease in absorbance of the chromogen released during hydrolysis of the peptide substrate is measured. The change in absorbance is proportional to the enzyme activity in the assay.

An alternate assay for ubiquitin hydrolase activity measures the hydrolysis of a ubiquitin precursor. The assay is performed at ambient temperature and contains an aliquot of PRTS and the appropriate substrate in a suitable buffer. For SEQ ID NO:1, chemically synthesized human ubiquitin-valine may be used as substrate. Cleavage of the C-terminal valine residue from the substrate is monitored by capillary electrophoresis (Franklin, K. et al. (1997) Anal. Biochem. 247:305-309).

Alternatively, the ubiquitin protease activity of SEQ ID NO:5 can be measured using the method of Sloper-Mould et al. ((1999) J. Biol. Chem. 274:26878-26884). Aliquots of PRTS are incubated with 5 μl [^{35}S]-labeled ubiquitin-GST fusion substrate for 1 hour at 37 °C in an appropriate buffer. Samples are resolved by electrophoresis on a 12% SDS-PAGE gel. Ubiquitin cleavage is monitored by fluorography of the gel.

Alternatively, the activity of SEQ ID NO:10, for example, can be measured by the method of Colige et al. (1999, J. Biol. Chem. 270:16724-16730). An aliquot of PRTS is incubated with amino procollagen type I substrate radioactively labeled only in the propeptide in a 250 μl reaction containing 50 mM sodium cacodylate, pH 7.5, 200 mM KCl, 2 mM $CaCl_2$, 2.5 mM NEM, 0.5 mM PMSF, and 0.02% Brij (standard assay buffer). After 16 h at 26 °C, the reaction is stopped by adding 50 μl of EDTA solution (0.2 M EDTA, pH 8, 0.5% SDS, 0.5 M DTT) and 300 μl of 99% ethanol. The samples are kept for 30 min at 4 °C and centrifuged for 30 min at 9500 g. Collagen and uncleaved

radioactive pN-collagen substrate are pelleted, whereas the freed amino propeptides remained in solution. An aliquot of the supernatant is assayed by liquid scintillation spectrometry.

In the alternative, an assay for protease activity takes advantage of fluorescence resonance energy transfer (FRET) that occurs when one donor and one acceptor fluorophore with an appropriate spectral overlap are in close proximity. A flexible peptide linker containing a cleavage site specific for PRTS is fused between a red-shifted variant (RSGFP4) and a blue variant (BFP5) of Green Fluorescent Protein. This fusion protein has spectral properties that suggest energy transfer is occurring from BFP5 to RSGFP4. When the fusion protein is incubated with PRTS, the substrate is cleaved, and the two fluorescent proteins dissociate. This is accompanied by a marked decrease in energy transfer which is quantified by comparing the emission spectra before and after the addition of PRTS (Mitra, R.D. et al. (1996) Gene 173:13-17). This assay can also be performed in living cells. In this case the fluorescent substrate protein is expressed constitutively in cells and PRTS is introduced on an inducible vector so that FRET can be monitored in the presence and absence of PRTS (Sagot, I. et al. (1999) FEBS Lett. 447:53-57).

In yet another alternative, an assay for PRTS dipeptidase activity measures the hydrolysis activity of PRTS on a variety of dipeptides such as leukotriene D4 (Kozak, E. and S. Tate (1982) J. Biol. Chem. 257:6322-6327), or hydrolysis of the beta-lactam ring of antibiotics such as penum and carbapenem (Campbell et al., (1984) J. Biol. Chem. 259:14586-14590).

XVIII. Identification of PRTS Substrates

Phage display libraries can be used to identify optimal substrate sequences for PRTS. A random hexamer followed by a linker and a known antibody epitope is cloned as an N-terminal extension of gene III in a filamentous phage library. Gene III codes for a coat protein, and the epitope will be displayed on the surface of each phage particle. The library is incubated with PRTS under proteolytic conditions so that the epitope will be removed if the hexamer codes for a PRTS cleavage site. An antibody that recognizes the epitope is added along with immobilized protein A. Uncleaved phage, which still bear the epitope, are removed by centrifugation. Phage in the supernatant are then amplified and undergo several more rounds of screening. Individual phage clones are then isolated and sequenced. Reaction kinetics for these peptide substrates can be studied using an assay in Example XVII, and an optimal cleavage sequence can be derived (Ke, S.H. et al. (1997) J. Biol. Chem. 272:16603-16609).

To screen for <u>in vivo</u> PRTS substrates, this method can be expanded to screen a cDNA expression library displayed on the surface of phage particles (T7SELECT 10-3 Phage display vector, Novagen, Madison WI) or yeast cells (pYD1 yeast display vector kit, Invitrogen, Carlsbad CA). In this case, entire cDNAs are fused between Gene III and the appropriate epitope.

XIX. Identification of PRTS Inhibitors

Compounds to be tested are arrayed in the wells of a multi-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. PRTS activity is measured for each well and the ability of each compound to inhibit PRTS activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance PRTS activity.

In the alternative, phage display libraries can be used to screen for peptide PRTS inhibitors. Candidates are found among peptides which bind tightly to a protease. In this case, multi-well plate wells are coated with PRTS and incubated with a random peptide phage display library or a cyclic peptide library (Koivunen, E. et al. (1999) Nat. Biotechnol. 17:768-774). Unbound phage are washed away and selected phage amplified and rescreened for several more rounds. Candidates are tested for PRTS inhibitory activity using an assay described in Example XVII.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
275791	1	275791CD1	22	275791CB1
1389845	2	1389845CD1	23	1389845CB1
1726609	3	1726609CD1	24	1726609CB1
4503848	4	4503848CD1	25	4503848CB1
5544089	5	5544089CD1	26	5544089CB1
7474081	6	7474081CD1	27	7474081CB1
5281209	7	5281209CD1	28	5281209CB1
2256251	8	2256251CD1	29	2256251CB1
7160544	9	7160544CD1	30	7160544CB1
7477386	10	7477386CD1	31	7477386CB1
7473089	11	7473089CD1	32	7473089CB1
7604035	12	7604035CD1	33	7604035CB1
3473847	13	3473847CD1	34	3473847CB1
3750004	14	3750004CD1	35	3750004CB1
4904126	15	4904126CD1	36	4904126CB1
71268415	16	71268415CD1	37	71268415CB1
7473301	17	7473301CD1	38	7473301CB1
7473308	18	7473308CD1	39	7473308CB1
7478021	19	7478021CD1	40	7478021CB1
4333459	20	4333459CD1	41	4333459CB1
6817347	21	6817347CD1	42	6817347CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	275791CD1	g9971757	1.00E-49	ubiquitin-specific processing protease [Homo sapiens]
2	1389845CD1	g1143194	1.30E-46	prostasin [Homo sapiens]
3	1726609CD1	g10719660	0	RNPEP-like protein [Homo sapiens] (Horikawa,Y. et al. (2000) Nat. Genet. 26:163-175.)
		g1754515	3.30E-96	aminopeptidase-B [Rattus norvegicus] (Prieto, I. et al. (1998) Horm. Metab. Res. 30:246-248.)
4	4503848CD1	g1783122	0	endopeptidase 24.16 type M1 [Sus scrofa]
5	5544089CD1	g5410230	5.20E-43	ubiquitin-specific protease 3 [Homo sapiens]
6	7474081CD1	g603903	2.90E-33	Trypsinogen [Gallus gallus]
7	5281209CD1	g11071729	0	putative dipeptidase [Homo sapiens]
		g219600	1.40E-88	dipeptidase precursor [Homo sapiens] (Satoh, S. et al. (1993) Biochim. Biophys. Acta 1172:181-183.)
8	2256251CD1	g6103629	3.90E-166	transmembrane tryptase [Homo sapiens] (Wong, G.W. et al. (1999) J. Biol. Chem. 274:30784-30793.)
9	7160544CD1	g11095188	0	dipeptidyl peptidase 8 [Homo sapiens] (Abbott,C.A. et al. (2000) Eur. J. Biochem. 267:6140-6150.)
		g1753197	6.80E-64	dipeptidyl peptidase IV [Stenotrophomonas maltophilia] (Mentlein, R. (1999) Regul. Pept. 85:9-24; Kahn , T. Int. J. Mol. Med (1999) 4:3-15.)
10	7477386CD1	g1865716	0	procollagen I N-proteinase [Bos taurus] (Colige, A. et al. (1999) Am. J. Hum. Genet. 65:308-317.)
11	7473089CD1	g7768706	3.60E-255	metalloprotease with thrombospondin type 1 motifs [Homo sapiens] (Vazquez, F. et al. (1999) J. Biol. Chem. 274:23349-23357.)
12	7604035CD1	g6164595	4.70E-68	Lacunin [Manduca sexta]

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
13	3473847CD1	g217172	9.20E-50	aqualysin precursor (aa 1 to 513) [Thermus aquaticus]
14	3750004CD1	g5923786	4.30E-51	zinc metalloprotease ADAMTS6 [Homo sapiens]
15	4904126CD1	g186286	3.90E-40	interleukin 1-beta convertase [Homo sapiens] (Cerretti, D.P. et al. (1992) Science 256:97-100.)
16	71268415CD1	g6651071	0	disintegrin and metalloproteinase domain 19 [Homo sapiens] (Inoue, D. et al. (1998) J. Biol. Chem. 273:4180-4187.)
17	7473301CD1	g6002714	5.10E-94	membrane-type serine protease 1 [Homo sapiens] (Takeuchi, T. et al. (1999) Proc. Natl. Acad. Sci. USA 96:11054-11061.)
18	7473308CD1	g1552517	6.60E-77	trypsinogen E [Homo sapiens]
19	7478021CD1	g3211705	5.60E-189	matrix metalloproteinase [Xenopus laevis] (Yang, M. (1997) J. Biol. Chem. 272:13527-13533.)
20	4333459CD1	g1754714	2.30E-67	Oviductin [Xenopus laevis] (Lindsay, L.L. et al. (1999) Biol. Reprod. 60:989-995.)
21	6817347CD1	g7673618	5.10E-283	ubiquitin specific protease [Mus musculus]

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	275791CD1	232	T15 T17 S23 S43 S71 T90 S93 S100 S107 S111 T122 T174 S190 T10 S141 S190 T213 T227	N98 N99	Ubiquitin carboxyl-terminal hydrolase family 2 signature 2 Uch_2_2: Y142-Y160	MOTIFS
					Ubiquitin carboxyl-terminal hydrolase family 2 signature 2 UCH-2: L138-H203	HMMER-PFAM
					Ubiquitin carboxyl-terminal hydrolase family 2 signature 2 BL00972: Y142-D166, K169-S190	BLIMPS-BLOCKS
2	1389845CD1	365	S120 S187 S225 S253 S82 T31 T37 T42 Y283		TRYPSIN DM00018|A57014|45-284: I123-Q314	BLAST_DOMO
					PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: I123-Q314	BLAST_PRODOM
					Serine proteases, trypsin family BL00134A: C148-C164	BLIMPS_BLOCKS
					Kringle domain proteins BL00021: C148-F165, V231-G252	BLIMPS_BLOCKS
					CHYMOTRYPSIN SERINE PROTEASE PR00722: G149-C164, G208-V222	BLIMPS_PRINTS
					V8 Serine proteases PR00839B: C148-F165	BLIMPS_PRINTS
					Serine proteases, trypsin family, active sites trypsin_his.prf: I140-P188	PROFILESCAN
					Trypsin: I123-Q314	HMMER_PFAM
					Trypsin_His: L159-C164	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	1726609CD1	416	S244 S283 S30 S370 S408 T389 T59 T78 T87	N203 N413 N57	do HYDROLASE; LEUKOTRIENE; A-4; ZINC; DM08707|P19602|7-609: M1-P354	BLAST_DOMO
					AMINOPEPTIDASE HYDROLASE METALLOPROTEASE ZINC N GLYCOPROTEIN PROTEIN TRANSMEMBRANE SIGNAL ANCHOR MEMBRANE PD001134: R4-S177	BLAST_PRODOM
					Neutral zinc metallopeptidase family BL00142: D41-F51	BLIMPS_BLOCKS
					MEMBRANE ALANYL DIPEPTIDASE PR00756: F14-L24, D41-T56, W60-Y72	BLIMPS_PRINTS
					signal_cleavage: M1-S34	SPScan
4	4503848CD1	714	S124 S140 S147 S179 S200 S206 S226 S333 S551 S556 S592 T114 T133 T244 T252 T270 T308 T318 T322 T376 T406 T432 T528 T585 T602 T69 Y175 Y249 Y505	N425 N485 N601	do ZINC; METALLOPEPTIDASE; NEUTRAL; OLIGO-PEPTIDASE DM01184|Q02038|36-702: A46-A713	BLAST_DOMO
					HYDROLASE METALLOPROTEASE ZINC OLIGOPEPTIDASE PRECURSOR MITOCHONDRIAL ENDOPEPTIDASE MITOCHONDRION TRANSIT PEPTIDE PD002945: W60-N529	BLAST_PRODOM
					transmembrane domain: L14-M34	HMMER
					Peptidase family M3 Peptidase_M3: C98-L711	HMMER_PFAM
					Neutral zinc metallopeptidase family BL00142: T504-H514	BLIMPS_BLOCKS
					Zinc_Protease: T504-M513	MOTIFS
					signal_cleavage: M1-G27	SPScan

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databas s
5	5544089CD1	367	S108 S161 S197 S203 S235 S266 S361 S49 T180 T263 T316 T331	N139 N142 N308	UBIQUITIN CARBOXYL-TERMINAL HYDROLASES FAMILY 2 DM00659|P40818|782-1103: L36-L328	BLAST_DOMO
					Ubiquitin carboxyl-terminal hydrolase family BL00972: Y74-L83, V120-C134, Y274-N298, G301-K322	BLIMPS_BLOCKS
					Ubiquitin carboxyl-terminal hydrolase family UCH-2: E270-Q332	HMMER_PFAM
					Ubiquitin carboxyl-terminal hydrolase family Uch_2_2: Y274-Y292	MOTIFS
					signal_cleavage: M1-S16	SPScan
6	7474081CD1	235	S134 S143 S159 S171 S226 T231	N90	TRYPSIN DM00018|S55065|26-244: A27-T231	BLAST_DOMO
					Serine proteases, trypsin family signature BL00134: C40-C56, V215-I228	BLIMPS_BLOCKS
					Type I fibronectin domain BL01253: C40-P53, I197-T231	BLIMPS_BLOCKS
					Kringle domain proteins BL00021: S159-Q164, C40-Y57	BLIMPS_BLOCKS
					CHYMOTRYPSIN SERINE PROTEASE PR00722A: V41-C56, A95-A109	BLIMPS_PRINTS
					Serine proteases, trypsin family, active site trypsin_his.prf: S35-T76	PROFILESCAN
					Trypsin: G42-V178, G216-I228	HMMER_PFAM
					Leucine_Zipper: L44-L65	MOTIFS
					signal_cleavage: M1-S19	SPScan

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	5281209CD1	488	S13 T74 S186 S233 T363 T456 T4 T34 T125 S170 S172 T178 S249 T337 S387 S389 S419 T447	N119 N184 N243 N334	Renal dipeptidase proteins BL00869:P92-L247, E280-R412, S415-N457	BLIMPS-BLOCKS
					DIPEPTIDASE MICROSOMAL PRECURSOR MDP HYDROLASE MICROSOME SIGNAL GPI-ANCHOR GLYCOPROTEIN ZINC PD005626:S143-E450	BLAST-PRODOM
					RENAL DIPEPTIDASE DM02775:T77-K410	BLAST-DOMO
					Renal dipeptidase:V195-R217	MOTIFS
					Renal dipeptase:A63-V475	HMMER-PFAM
					Signal peptide:M1-A36	HMMER
					Signal cleavage:M1-A36	SPSCAN
8	2256251CD1	346	S203 S210 S266 S45 S79 T131 T147 T216	N110	TRYPSIN DM00018 | P15944 | 31-270: I63-I294	BLAST_DOMO
					PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: L156-I290, I63-F178, N288-F314, P274-P305	BLAST_PRODOM
					Serine proteases, trypsin family BL00134: C88-C104, D241-I264, P277-I290	BLIMPS_BLOCKS
					Type I fibronectin domain BL01253: C88-A101, V158-E194, G240-C253, W259-H293	BLIMPS_BLOCKS
					Kringle domain proteins BL00021: C88-F105, V169-G190, G249-I290	BLIMPS_BLOCKS
					CHYMOTRYPSIN SERINE PROTEASE PR00722: G89-C104, G146-V160, G240-V252	BLIMPS_PRINTS
					transmembrane domain: P308-L328	HMMER
					trypsin: I63-I290	HMMER_PFAM
					Trypsin_His L99-C104	MOTIFS
					Trypsin family serine protease active sites trypsin_his.prf: L80-H128 trypsin_ser.prf: L229-R273	PROFILESCAN
					signal_cleavage: M1-S45	SPSCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	7160544CD1	882	S115 S133 S293 S312 S412 S443 S479 S530 S587 S588 S723 S80 S850 T227 T234 T307 T326 T499 T52 T551 T594 T603 T615 T776 Y315 Y36 Y55 Y555 Y844		PROLYL ENDOPEPTIDASE FAMILY SERINE DM02461| P27487|192-765: F488-E870, G251-E370	BLAST_DOMO
					DIPEPTIDYL IV HYDROLASE PROTEASE SERINE PEPTIDASE DIPEPTIDASE TRANSMEMBRANE GLYCOPROTEIN PROTEIN PD003048: L744-E870	BLAST_PRODOM
					DIPEPTIDYL IV HYDROLASE PROTEASE SERINE PEPTIDASE DIPEPTIDASE TRANSMEMBRANE GLYCOPROTEIN PROTEIN PD003086: Y423-V661, I212-T326	BLAST_PRODOM
					Prolyl endopeptidase family BL00708: G501-I513, Q714-L744	BLIMPS_BLOCKS
					Dipeptidyl peptidase IV PF00930: I498-R508, F756-P783, R808-L828	BLIMPS_PFAM
					PROLYL OLIGOPEPTIDASE SERINE PROTEASE PR00862: P647-F665, G737-R757	BLIMPS_PRINTS
					Prolyl oligopeptidase family Peptidase_S9: R672-L744	HMMER_PFAM
					Dipeptidyl peptidase IV DPPIV_N_term: M88-N663	HMMER_PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10	7477386CD1	1189	S132 S169 S200 S32 S323 S350 S445 S480 S511 S626 S675 S699 S798 S1064 T247 T362 T521 T612 T718 T777 T946 T986 T1104 Y552	N109 N478 N944	do ZINC; METALLOPEPTIDASE; NEUTRAL; ATROLYSIN DM00368|Q05910|189-395:I261-P463	BLAST_DOMO
					THROMBOSPONDIN TYPE 1 REPEAT DM00275|P07996|477-540: D555-C604	BLAST_DOMO
					PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE C02B4.1 A DISINTEGRIN METALLOPROTEASE PD013511: L474-E549	BLAST_PRODOM
					PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654: Q647-C716	BLAST_PRODOM
					PROCOLLAGEN C37C3.6 SERINE PROTEASE INHIBITOR PD007018: W854-Q974, W914-C1029, W558-K623	BLAST_PRODOM
					METALLOPROTEASE PRECURSOR HYDROLASE SIGNAL ZINC VENOM CELL PROTEIN TRANSMEMBRANE ADHESION PD000791:P256-P463	BLAST_PRODOM
					Neutral zinc metallopeptidase BL00142: T398-N408	BLIMPS_BLOCKS
					signal_peptide: M1-A22	HMMER
					Reprolysin family propeptide Pep_M12B_propep: R120-V240	HMMER_PFAM
					Reprolysin (M12B) family zinc metalloprotease Reprolysin: I261-P463	HMMER_PFAM
					Thrombospondin type 1 domain tsp_1: A973-C1024, S559-C609, Y852-C909, W914-C971	HMMER_PFAM
					signal_cleavage: M1-G23	SPSCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	7473089CD1	952	S19 S203 S207 S303 S346 S432 S492 S575 S578 S611 S666 S682 S708 S745 S919 T171 T288 T317 T325 T337 T359 T471 T594 T687 T765	N141 N591 N623 N680	do ZINC; METALLOPEPTIDASE; NEUTRAL; ATROLYSIN DM00368|JC2550|1-201:R218-P427	BLAST_DOMO
					PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD014161: K684-E804	BLAST_PRODOM
					PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654: V610-C683	BLAST_PRODOM
					METALLOPROTEASE PRECURSOR HYDROLASE SIGNAL ZINC VENOM CELL PROTEIN TRANSMEMBRANE ADHESION PD000791: V214-P427	BLAST_PRODOM
					PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE C02B4.1 A DISINTEGRIN METALLOPROTEASE PD013511: L437-V505	BLAST_PRODOM
					Neutral zinc metallopeptidase BL00142: T358-N368	BLIMPS_BLOCKS
					signal_peptide: M1-G18	HMMER
					Reprolysin family propeptide Pep_M12B_propep: H67-N181	HMMER_PFAM
					Reprolysin (M12B) family zinc metalloprotease Reprolysin: R218-P427	HMMER_PFAM
					Thrombospondin type 1 domain tsp_1: A520-C570, W845-C896, W899-C952	HMMER_PFAM
					Zinc_Protease: T358-F367	MOTIFS
					Spscan signal_cleavage: M1-G17	SPSCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	7604035CD1	898	S187 S188 S258 S268 S285 S415 S467 S547 S696 S796 S819 S851 S892 T106 T198 T35 T434 T483 T492 T5	N3 N490 N773	PROCOLLAGEN C37C3.6 SERINE PROTEASE INHIBITOR ALTERNATIVE PD007018: Y726-C841, W786-A874, Y667-C778, W50-Q72, S368-Q383	BLAST_PRODOM
					PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654: Q416-C484	BLAST_PRODOM
					PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD014161: R485-I599	BLAST_PRODOM
					signal_peptide: M1-D24	HMMER
					Thrombospondin type 1 domain tsp_1: G48-R87, W727-C783, E787-C841	HMMER_PFAM
					signal_cleavage: M1-D24	SPSCAN
13	3473847CD1	631	S117 S160 S174 S185 S188 S268 S28 S30 S358 S431 S503 S605 T142 T33 T346 T512 T606	N472	Subtilase family Peptidase_S8: S86-N364	HMMER_PFAM
					SERINE PROTEASES, SUBTILASE FAMILY, HISTIDINE DM00108|P80146|150-377: G116-T346	BLAST_DOMO
					Serine proteases, subtilase family BL00136: L123-I135, D163-G175, G323-G333	BLIMPS_BLOCKS
					SUBTILISIN SERINE PROTEASE FAMILY PR00723: G116-I135, K161-S174, S322-M338	BLIMPS_PRINTS
					Serine proteases, subtilase family, active site subtilase_ser.prf: A302-Q352	PROFILESCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	3750004CD1	470	S454 S51 S54 T104 T276 T386 T464	N182 N203	Thrombospondin type 1 domain tsp_1: T34-C81	HMMER_PFAM
					PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654: Q119-C185	BLAST_PRODOM
					signal_peptide: M1-G29	HMMER
					signal_cleavage: M1-G24	SPScan
15	4904126CD1	110	S16 S36 T100 T49	N47	Caspase recruitment domain CARD: A2-A91	HMMER_PFAM
					INTERLEUKIN-1 BETA CONVERTING ENZYME FAMILY HISTIDINE DM07463|P29466|1-122: M1-S110	BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	71268415CD1	879	S132 S14 S208 S288 S571 S711 S747 S754 S755 S827 T106 T118 T29 T30 T373 T412 T42 T428 T444 T55 T688 Y167 Y39	N368 N371 N569 N68	Reprolysin family propeptide Pep_M12B_propep: R8-K119	HMMER_PFAM
					Reprolysin (M12B) family zinc metalloprotease Reprolysin: K134-P332	HMMER_PFAM
					Disintegrin: E349-Q424	HMMER_PFAM
					do ZINC; METALLOPEPTIDASE; NEUTRAL; ATROLYSIN; DM00368|S60257|204-414: K126-D333	BLAST_DOMO
					do ZINC; REGULATED; EPIDIDYMAL; NEUTRAL; DM00591|S60257|492-628: F410-L549	BLAST_DOMO
					MELTRIN, BETA METALLOPROTEASE DISINTEGRIN BETA INTEGRIN PROTEASE METALLOPROTEASE PD105322: P620-G812	BLAST_PRODOM
					METALLOPROTEASE PRECURSOR HYDROLASE SIGNAL ZINC VENOM CELL PROTEIN TRANSMEMBRANE ADHESION PD000791: K134-P332	BLAST_PRODOM
					MELTRIN, BETA METALLOPROTEASE DISINTEGRIN MELTRIN BETA INTEGRIN PROTEASE METALLOPROTEASE PD171676: K495-C567	BLAST_PRODOM
					CELL ADHESION PLATELET BLOOD COAGULATION VENOM DISINTEGRIN METALLOPROTEASE PRECURSOR SIGNAL PD000664: E349-Y423	BLAST_PRODOM
					Neutral zinc metalloprotease BL00142: T266-G276	BLIMPS_BLOCKS
					DISINTEGRIN SIGNATURE PR00289: C380-R399, E409-N421	BLIMPS_PRINTS
					NEPRILYSIN METALLOPROTEASE PR00786C: N259-F275	BLIMPS_PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Disintegrins signature disintegrins.prf: E360-P419	PROFILESCAN
					Neutral zinc metallopeptidases, zinc-binding region signature zinc_protease.prf: S249-G301	PROFILESCAN
					transmembrane domain: V624-Y645	HMMER
					Zn binding region Zinc_Protease: T266-F275	MOTIFS
17	7473301CD1	850	S100 S275 S295 S358 S429 S448 S470 S474 S495 S536 S596 S64 S787 S802 S807 T117 T250 T312 T348 T382 T404 T426 T570 T714 T777	N19 N210 N422 N486 N533 N559 N568	TRYPSIN FAMILY SERINE PROTEASE trypsin: I613-I842	HMMER_PFAM
					Low-density lipoprotein receptor domain ldl_recept_a: Q489-S527, P530-Q562, I564-C603	HMMER_PFAM
					TRYPSIN DM00018|P98072|800-1033: R612-V846	BLAST_DOMO
					PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: Q675-I842, I613-G809	BLAST_PRODOM
					Serine proteases, trypsin family BL00134: C638-C654, D791-T814, P829-I842	BLIMPS_BLOCKS
					Type I fibronectin domain BL01253: C638-A651, R790-C803, W811-Y845	BLIMPS_BLOCKS
					Kringle domain proteins BL00021: I722-G743, L801-I842, C638-F655	BLIMPS_BLOCKS
					LOW DENSITY LIPOPROTEIN RECEPTOR DOMAIN PR00261: G501-E522	BLIMPS_PRINTS
					CHYMOTRYPSIN SERINE PROTEASE PR00722: G639-C654, T697-W711, R790-S802	BLIMPS_PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Trypsin family serine protease active sites trypsin_his.prf: L630-K679 trypsin_ser.prf: I776-R825	PROFILESCAN
					transmembrane domain: I77-L95	HMMER
					Trypsin family serine protease active sites Trypsin_His L649-C654 Trypsin_Ser D791-S802	MOTIFS
18	7473308CD1	254	S136 S14 S153 S195 S227 T230 T249		TRYPSIN FAMILY SERINE PROTEASE trypsin: I21-Q183	HMMER_PFAM
					CHYMOTRYPSIN SERINE PROTEASE FAMILY PR00722B: T89-A103	BLIMPS_PRINTS
					TRYPSIN DM00018|P07478|24-242: I21-Q183	BLAST_DOMO
					PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: G23-Q183	BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
19	7478021CD1	568	S142 S145 S153 S172 S177 S190 S244 S316 S34 S420 S448 S552 T209 T22 T293 T334 T401 T427 T489 T79 Y509	N371	Matrixin Peptidase_M10: F56-T266	HMMER_PFAM
					Hemopexin domain: F332-T390, I393-S448, L450-Q498, I505-K548	HMMER_PFAM
					MATRIXINS CYSTEINE SWITCH DM00558|P22757|100-337: A184-T334, P76-P124	BLAST_DOMO
					MATRIXINS CYSTEINE SWITCH DM00558|P08254|29-274: Q158-T334, L85-M122	BLAST_DOMO
					MATRIX METALLOPROTEINASE PD168921: S327-N392	BLAST_PRODOM
					MATRIX METALLOPROTEINASE PD169970: A494-M568	BLAST_PRODOM
					MATRIX PRECURSOR METALLOPROTEASE HYDROLASE ZINC ZYMOGEN CALCIUM COLLAGEN DEGRADATION SIGNAL PD000673: F171-T266, P73-M122	BLAST_PRODOM
					Matrixins cysteine switch BL00546: F92-D121, V224-P267, G273-Y304, L313-G326, F443-Y455, F409-E428	BLIMPS_BLOCKS
					Hemopexin domain protein BL00024: M112-M122, G273-Y304, L313-G326, F443-Y455, Y408-D419	BLIMPS_BLOCKS
					MATRIXIN SIGNATURE PR00138: M112-P125, E198-F213, V224-W252, V279-Y304, L313-G326	BLIMPS_PRINTS
					Matrixins cysteine switch cysteine_switch.prf: A95-M204	PROFILESCAN
					Neutral zinc metallopeptidases, zinc-binding region signature zinc_protease.prf: D256-E312	PROFILESCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Hemopexin domain signature hemopexin.prf: F409-R477	PROFILESCAN
					signal_peptide: M1-P21	HMMER
					Zn binding region Zinc_Protease: V279-L288	MOTIFS
					signal_cleavage: M1-P24	SPScan
20	4333459CD1	306	S117 S138 S2 S223 S60 S72 T110 T139 T207 T217	N108	TRYPSIN FAMILY SERINE PROTEASE trypsin:I56-I298	HMMER_PFAM
					TRYPSIN DM00018|Q05319|543-784: I56-I302	BLAST_DOMO
					PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: S117-I298, I56-G192	BLAST_PRODOM
					Serine proteases, trypsin family BL00134: C81-C97, D238-G261, P285-I298	BLIMPS_BLOCKS
					Type I fibronectin domain BL01253: C81-A94, G154-E190, R237-C250	BLIMPS_BLOCKS
					Kringle domain proteins BL00021: C81-I98, I165-G186, S247-F288	BLIMPS_BLOCKS
					CHYMOTRYPSIN SERINE PROTEASE PR00722: G82-C97, P142-F156, R237-M249	BLIMPS_PRINTS
					Trypsin family serine protease active sites trypsin_his.prf: L73-G122 trypsin_ser.prf: K225-R271	PROFILESCAN
					Trypsin family serine protease active sites Trypsin_His: I92-C97 Trypsin_Ser: D238-M249	MOTIFS
					signal_cleavage: M1-S26	SPScan

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
21	6817347CD1	953	S102 S114 S150 S172 S369 S429 S47 S623 S794 S804 S808 S831 S856 S919 S942 T289 T42 T455 T488 T544 T567 T568 T585 T59 T736 T777 T786 T839 Y929	N95	Ubiquitin carboxyl-terminal hydrolase family 1 UCH-1: R593-D624	HMMER_PFAM
					Ubiquitin carboxyl-terminal hydrolase family 2 UCH-2: N875-K935	HMMER_PFAM
					UBIQUITIN CARBOXYL-TERMINAL HYDROLASES FAMILY 2 DM00659|P40818|782-1103: T777-L931, L598-H709, I713-T753, V101-L128	BLAST_DOMO
					PROTEASE UBIQUITIN HYDROLASE UBIQUITIN SPECIFIC ENZYME DEUBIQUITINATING C-TERMINAL THIOLESTERASE PROCESSING CONJUGATION PD017412: T777-E859	BLAST_PRODOM
					Ubiquitin carboxyl-terminal hydrolase family 2 BL00972: G594-L611, Y675-L684, I714-C728, K878-H902, K904-D925	BLIMPS_BLOCKS
					Ubiquitin carboxyl-terminal hydrolase family 2 signature 1 Uch_2_1: G594-Q609	MOTIFS
					Ubiquitin carboxyl-terminal hydrolase family 2 signature 2 Uch_2_2: Y879-Y896	MOTIFS

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
22	275791CB1	2204	1168-1197, 1503-1522, 1-281, 1716-1738	6456514H1 (COLNDIC01)	890	1510
				7246467T8 (PROSTMY01)	692	1390
				4943009F8 (BRAIFEN05)	1319	1877
				55047202J1	1	811
				6053385H1 (BRABDIR03)	1636	2204
23	1389845CB1	2036	1-392, 1468-1491, 1334-1400, 1974-2036	FL1389845_00001	6	2036
				1389845H1 (EOSINOT01)	1	244
24	1726609CB1	2185	1-44, 1804-2185	71762189V1	1	662
				5426388F9 (PROSTMT07)	1370	1992
				71053940V1	744	1373
				71041539V1	1954	2185
				5968441H1 (BRAZNOT01)	1326	1824
				6756865J1 (SINTFER02)	642	1316
25	4503848CB1	3486	1-1330	2053131H1 (BEPINOT01)	2885	3136
				6440674H1 (BRAENOT02)	2576	3095
				g5745066	1831	2254
				7191212H2 (BRATDIC01)	2295	2873
				5960039H1 (BRATNOT05)	1229	1797
				GBI.g7710158_edit	1	2015
				60200050D1	825	1146
				5969176H1 (BRAZNOT01)	1677	2104
				5649471H1 (BRAITUT23)	3157	3486
				2232143F6 (PROSNOT16)	2235	2670
				3022114H1 (PROSDIN01)	3116	3402
				60220456D1	1085	1462

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
26	5544089CB1	2847	1260-1631, 2532-2847, 408-914	55051688J1	810	1767
				2344450F6 (TESTTUT02)	2195	2847
				7658834H1 (OVARNOE02)	2130	2642
				71763578V1	1562	2266
				6576463H1 (COLHTUS02)	1079	1838
				55051680J1.comp	1	958
27	7474081CB1	890	1-21	g2103202	1	493
				g2142177	397	890
28	5281209CB1	1577	1-629	FL5281209_g7712102_000004_g436191	1	1467
				g3644494	1155	1577
				3142983R6 (HNT2AZS07)	1045	1576
29	2256251CB1	1958	1-399, 896-935	3220504T6 (COLNNON03)	1611	1958
				g4264312	400	848
				FL2256251_g7708357_000002_g6103629	910	1830
				2256251R6 (OVARTUT01)	408	1003
30	7160544CB1	3106	1-540, 1166-1428	6471337H1 (PLACFEB01)	1654	2316
				6854305H1 (BRAIFEN08)	895	1561
				4443368H1 (SINDNOT01)	1332	1585
				6894004J1 (BRAITDR03)	470	1077
				7655990H1 (UTREDME06)	324	822
				7745974H1 (ADRETUE04)	2378	3052
				7160544H1 (HNT2TXC01)	1	427
				70490289V1	2636	3106
				70748463V1	2269	2891
				7745974J1 (ADRETUE04)	1566	2153

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
31	7477386CB1	3567	1-971, 1953-2846, 3243-3567	GBI:g6682143_000029_edit.20231-20345	1495	1608
				GBI:g6682143_000023.comp_edit.11365-11445	1	81
				GBI:g6682143_000027_edit.14110-14265	523	678
				GBI:g6682143_000029_edit.35032-35202	2272	2436
				GBI:g6682143_000029_edit.13651-13803	958	1110
				GBI:g6682143_000019_edit.3461-3655	679	873
				GBI:g6682143_000029_edit.41513-41644	2944	3075
				GBI:g6682143_000029_edit.43912-44404	3076	3567
				GBI:g6682143_000029_edit.12846-12930	874	957
				GBI:g6682143_000029_edit.15804-15912	1111	1218
				GBI:g6682143_000023.comp_edit.9253-9693	82	522
				GBI:g6682143_000029_edit.27621-27746	2068	2193
				GBI:g6682143_000029_edit.18722-18853	1363	1494
				GBI:g6682143_000029_edit.17438-17581	1219	1362
				GBI:g6682143_000029_edit.35651-35783	2608	2739

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
22	275791CB1	2204	1168-1197, 1503-1522, 1-281, 1716-1738	6456514H1 (COLNDIC01)	890	1510
				7246467T8 (PROSTMY01)	692	1390
				4943009F8 (BRAIFEN05)	1319	1877
				55047202J1	1	811
				6053385H1 (BRABDIR03)	1636	2204
23	1389845CB1	2036	1-392, 1468-1491, 1334-1400, 1974-2036	FL1389845_00001	6	2036
				1389845H1 (EOSINOT01)	1	244
24	1726609CB1	2185	1-44, 1804-2185	71762189V1	1	662
				5426388F9 (PROSTMT07)	1370	1992
				71053940V1	744	1373
				71041539V1	1954	2185
				5968441H1 (BRAZNOT01)	1326	1824
				6756865J1 (SINTFER02)	642	1316
25	4503848CB1	3486	1-1330	2053131H1 (BEPINOT01)	2885	3136
				6440674H1 (BRAENOT02)	2576	3095
				g5745066	1831	2254
				7191212H2 (BRATDIC01)	2295	2873
				5960039H1 (BRATNOT05)	1229	1797
				GBI.g7710158_edit	1	2015
				60200050D1	825	1146
				5969176H1 (BRAZNOT01)	1677	2104
				5649471H1 (BRAITUT23)	3157	3486
				2232143F6 (PROSNOT16)	2235	2670
				3022114H1 (PROSDIN01)	3116	3402
				60220456D1	1085	1462

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
26	5544089CB1	2847	1260-1631, 2532-2847, 408-914	55051688J1	810	1767
				2344450F6 (TESTTUT02)	2195	2847
				7658834H1 (OVARNOE02)	2130	2642
				71763578V1	1562	2266
				6576463H1 (COLHTUS02)	1079	1838
				55051680J1.comp	1	958
27	7474081CB1	890	1-21	g2103202	1	493
				g2142177	397	890
28	5281209CB1	1577	1-629	FL5281209_g7712102_000004_g436191	1	1467
				g3644494	1155	1577
				3142983R6 (HNT2AZS07)	1045	1576
29	2256251CB1	1958	1-399, 896-935	3220504T6 (COLNNON03)	1611	1958
				g4264312	400	848
				FL2256251_g7708357_000002_g6103629	910	1830
				2256251R6 (OVARTUT01)	408	1003
30	7160544CB1	3106	1-540, 1166-1428	6471337H1 (PLACFEB01)	1654	2316
				6854305H1 (BRAIFEN08)	895	1561
				4443368H1 (SINDNOT01)	1332	1585
				6894004J1 (BRAITDR03)	470	1077
				7655990H1 (UTREDME06)	324	822
				7745974H1 (ADRETUE04)	2378	3052
				7160544H1 (HNT2TXC01)	1	427
				70490289V1	2636	3106
				70748463V1	2269	2891
				7745974J1 (ADRETUE04)	1566	2153

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
31	7477386CB1	3567	1-971, 1953-2846, 3243-3567	GBI:g6682143_000029_edit.20231-20345	1495	1608
				GBI:g6682143_000023.comp_edit.11365-11445	1	81
				GBI:g6682143_000027_edit.14110-14265	523	678
				GBI:g6682143_000029_edit.35032-35202	2272	2436
				GBI:g6682143_000029_edit.13651-13803	958	1110
				GBI:g6682143_000019_edit.3461-3655	679	873
				GBI:g6682143_000029_edit.41513-41644	2944	3075
				GBI:g6682143_000029_edit.43912-44404	3076	3567
				GBI:g6682143_000029_edit.12846-12930	874	957
				GBI:g6682143_000029_edit.15804-15912	1111	1218
				GBI:g6682143_000023.comp_edit.9253-9693	82	522
				GBI:g6682143_000029_edit.27621-27746	2068	2193
				GBI:g6682143_000029_edit.18722-18853	1363	1494
				GBI:g6682143_000029_edit.17438-17581	1219	1362
				GBI:g6682143_000029_edit.35651-35783	2608	2739

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
				GBI:g6682143_000029_edit.27099-27237	1759	1932
				GBI:g6682143_000029_edit_24540-24713	1609	1758
				GBI:g6682143_000029_edit.37355-37558	2740	2943
32	7473089CB1	2930	1-632, 1082-1138, 2453-2555, 1373-1615, 1716-1740	GBI:g7387384_000011.comp_edit.13491-13892	2529	2930
				GBI:g7387384_000010_edit.2924-3211	1335	1619
				GBI:g7387384_000010_edit.13479-13850	2157	2528
				GBI:g7387384_000010_edit.11350-11535	1794	1979
				7988641H1 (UTRSTUC01)	1064	1587
				GBI:g7387384_000010_edit.9694-9867	1620	1793
				GBI:g7387384_000012.comp.edit_9639-10595	75	1031
				GBI:g7387384_000010_edit.1917-2074	1032	1163
				GBI:g7387384_000010_edit.2514-2684	1164	1334
				7631548J1 (BRAFTUE03)	21	619
				GBI:g7387384_000010_edit.11639-11815	1980	2156
				GBI:g7387384_edit	1	2930

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
33	7604035CB1	4230	4185-4230, 894-2774	6254235H1 (LUNPTUT02)	3308	3897
				6213818H1 (MUSCDIT06)	3900	4230
				6314348H1 (NERDTDN03)	3426	3994
				7195502H1 (LUNGFER04)	2758	3376
				6800634J1 (COLENOR03)	2678	3323
				8113675H1 (OSTEUNC01)	1661	2049
				6804059H1 (COLENOR03)	478	1050
				7750274H1 (HEAONOE01)	1995	2503
				3843227F6 (DENDNOT01)	2094	2707
				7632961H1 (BLADTUE01)	1418	2024
				55097977J1	688	1535
				7716357J1 (SINTFEE02)	1	678
34	3473847CB1	3699	1-2631	71906145V1	1340	2118
				7101935F8 (BRAWTDR02)	166	760
				70857826V1	2757	3441
				70855756V1	2650	3239
				820867R1 (KERANOT02)	2166	2732
				8055446J1 (ESOGTUE01)	579	1013
				70857738V1	3156	3699
				70858612V1	1998	2671
				GNN.g7208751_000002_002.edit	555	1850
				7101935R8 (BRAWTDR02)	1	473

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
35	3750004CB1	2410	1-264, 2116-2410, 1057-1167, 1590-1649	g7712021_edit	1	246
				7680089J1 (BRAFTUE01)	1327	1911
				6804411H1 (COLENOR03)	1088	1618
				71909368V1	536	968
				g1187194	1655	2127
				g2241985	706	1144
				6314962H1 (NERDTDN03)	973	1135
				6823371J1 (SINTNOR01)	65	855
				7655009J1 (UTREDME06)	1407	1990
				g1272147	1754	2410
36	4904126CB1	549		71620969V1	1	549
37	71268415CB1	2755	1-1097, 2326-2755	7715927J1 (SINTFEE02)	590	1340
				7372052H2 (BRAIFEE04)	2044	2514
				g6651070_CD	102	2755
				7723192J2 (THYRDIE01)	905	1500
				GBI:g7709257_000011.edit	1	139
				8037549H1 (SMCRUNE01)	206	819
				7720289J1 (THYRDIE01)	1596	2263
				8037549J1 (SMCRUNE01)	1456	2120

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
38	7473301CB1	2553	1-2394	GBI.g7272157_000017.edit	2001	2553
				71704195V1	1713	2016
				5544473H1 (TESTNOC01)	622	680
				GNN.g7272157_000017_002.edit	1688	2382
				5544473T8 (TESTNOC01)	2246	2550
				71703469V1	1163	1746
				GNN.g8571511_000004_002.edit	981	1468
				GNN.g6624046_000008_004	1	1111
39	7473308CB1	1041	826-1041, 1-299	GNN.g1552511_035	1	1041
40	7478021CB1	1707	1-1188	g8176728_edit	979	1083
				g7684439_edit	1	978
				g7684439_edit_2	1084	1707
41	4333459CB1	1262	1-1262	71571956V1	704	1262
				5634861R8 (PLACFER01)	1	266
				71571988V1	256	937
				71573159V1	247	928
42	6817347CB1	3067	1-2270	55022864H1	2314	3067
				55022792H2	1392	2091
				55022814H1	2080	2886
				55022795J2	2044	2726
				GNN.g7417337_004.edit	1	3067

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
22	275791CB1	TESTNOT03
23	1389845CB1	EOSITXT01
24	1726609CB1	BRAITUT02
25	4503848CB1	PROSNOT16
26	5544089CB1	BRAIFEC01
28	5281209CB1	HNT2AZS07
29	2256251CB1	OVARTUT01
30	7160544CB1	BRAFNOT02
32	7473089CB1	UTRSTUC01
33	7604035CB1	PLACNOR01
34	3473847CB1	KERANOT02
35	3750004CB1	BRAFTUE01
36	4904126CB1	TLYMNOT08
37	71268415CB1	THYRDIE01
38	7473301CB1	TESTNOC01
41	4333459CB1	KIDCTMT01
42	6817347CB1	ADRETUR01

Table 6

Library	Vector	Library Description
ADRETUR01	PCDNA2.1	This random primed library was constructed using RNA isolated from left upper pole, adrenal gland tumor tissue removed from a 52-year-old Caucasian male during nephroureterectomy and local destruction of renal lesion. Pathology indicated grade 3 adrenal cortical carcinoma forming a mass that infiltrated almost the whole adrenal parenchyma and extended to adjacent adipose tissue. A metastatic tumor nodule was identified in the hilar region. The renal vein was infiltrated by tumor and the neoplastic process was present at the resection margin of the renal vein. Fragments of adrenal cortical carcinoma and thrombus were found in the inferior vena cava. Patient history included abnormal weight loss. Family history included skin cancer, type I diabetes, and neurotic depression.
BRAFNOT02	pINCY	Library was constructed using RNA isolated from superior frontal cortex tissue removed from a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Microscopically, the cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. In addition, scattered throughout the cerebral cortex, there were multiple small microscopic areas of cavitation with surrounding gliosis. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly, and an enlarged spleen and liver.
BRAFTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from brain tumor tissue removed from the frontal lobe of a 58-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated a grade 2 metastatic hypernephroma. The patient presented with migraine headache. The patient developed a cerebral hemorrhage and pulmonary edema, and died during this hospitalization. Patient history included a grade 2 renal cell carcinoma, insomnia, and chronic airway obstruction. Previous surgeries included a nephroureterectomy. Patient medications included Decadron and Dilantin. Family history included a malignant neoplasm of the kidney in the father.
BRAIFEC01	pINCY	This large size-fractionated library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.

Table 6 (cont.)

Library	Vector	Library Description
BRAITUT02	PSPORT1	Library was constructed using RNA isolated from brain tumor tissue removed from the frontal lobe of a 58-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated a grade 2 metastatic hypernephroma. Patient history included a grade 2 renal cell carcinoma, insomnia, and chronic airway obstruction. Family history included a malignant neoplasm of the kidney.
EOSITXT01	pINCY	Library was constructed using RNA isolated from eosinophils stimulated with IL-5.
HNT2AZS07	PSPORT1	This subtracted library was constructed from RNA isolated from an hNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor) treated for three days with 0.35 micromolar AZ. The hybridization probe for subtraction was derived from a similarly constructed library from untreated hNT2 cells. 3.08M clones from the AZ-treated library were subjected to three rounds of subtractive hybridization with 3.04M clones from the untreated library. Subtractive hybridization conditions were based on the methodologies of Swaroop et al. (NAR (1991) 19:1954) and Bonaldo et al. (Genome Research (1996) 6:791).
KERANOT02	PSPORT1	Library was constructed using RNA isolated from epidermal breast keratinocytes (NHEK). NHEK (Clontech #CC-2501) is human breast keratinocyte cell line derived from a 30-year-old black female during breast-reduction surgery.
KIDCTMT01	pINCY	Library was constructed using RNA isolated from kidney cortex tissue removed from a 65-year-old male during nephroureterectomy. Pathology for the associated tumor tissue indicated grade 3 renal cell carcinoma within the mid-portion of the kidney and the renal capsule.
OVARTUT01	PSPORT1	Library was constructed using RNA isolated from ovarian tumor tissue removed from a 43-year-old Caucasian female during removal of the fallopian tubes and ovaries. Pathology indicated grade 2 mucinous cystadenocarcinoma involving the entire left ovary. Patient history included mitral valve disorder, pneumonia, and viral hepatitis. Family history included atherosclerotic coronary artery disease, pancreatic cancer, stress reaction, cerebrovascular disease, breast cancer, and uterine cancer.

Table 6 (cont.)

Library	Vector	Library Description
PLACNOR01	PCDNA2.1	This random primed library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from placental tissue removed from a Caucasian fetus (donor A), who died after 16 weeks' gestation from fetal demise and hydrocephalus and from placental tissue removed from a Caucasian male fetus (donor B), who died after 18 weeks' gestation from fetal demise. Patient history for donor A included umbilical cord wrapped around the head (3 times) and the shoulders (1 time). Serology was positive for anti-CMV and remaining serologi s were negative. Family history included multiple pregnancies and live births, and an abortion in the mother. Serology was negative for donor B.
PROSNOT16	pINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 68-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). During this hospitalization, the patient was diagnosed with myasthenia gravis. Patient history included osteoarthritis, and type II diabetes. Family history included benign hypertension, acute myocardial infarction, hyperlipidemia, and arteriosclerotic coronary artery disease.
TESTNOC01	PBLUESCRIPT	This large size fractionated library was constructed using RNA isolated from testicular tissue removed from a pool of eleven, 10 to 61-year-old Caucasian males.
TESTNOT03	PBLUESCRIPT	Library was constructed using RNA isolated from testicular tissue removed from a 37-year-old Caucasian male, who died from liver disease. Patient history included cirrhosis, jaundice, and liver failure.
THYRDIE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from diseased thyroid tissue removed from a 22-year-old Caucasian female during closed thyroid biopsy, partial thyroidectomy, and regional lymph node excision. Pathology indicated adenomatous hyperplasia. The patient presented with malignant neoplasm of the thyroid. Patient history included normal delivery, alcohol abuse, and tobacco abuse. Previous surgeries included myringotomy. Patient medications included an unspecified type of birth control pills. Family history included hyperlipidemia and depressive disorder in the mother; and benign hypertension, congestive heart failure, and chronic leukemia in the grandparent(s).

Table 6 (cont.)

Library	Vector	Library Description
TLYMNOT08	pINCY	The library was constructed using RNA isolated from anergicallogenic T-lymphocyte tissue removed from an adult (40-50-year-old) Caucasian male.The cells were incubated for 3 days in the presence of 1 microgram/ml OKT3 mAb and 5% human serum.
UTRSTUC01	PSPORT1	This large size fractionated library was constructed using pooled cDNA from two donors. cDNA was generated using mRNA isolated from uterus tumor tissue remov d from a 37-year-old Black female (donor A) during myomectomy, dilation and curettage, right fimbrial region biopsy, and incidental appendectomy; and from endometrial tumor tissue removed from a 49-year-old Caucasian female (donor B) during vaginal hysterectomy and bilateral salpingo-oophorectomy. For donor A, pathology indicated multiple uterine leiomyomata. A fimbrial cyst was identified. The endometrium was in secretory phase with hormonal effect. The patient presented with deficiency anemia, an umbilical hernia, and premenopausal menorrhagia. Patient history included premenopausal menorrhagia and sarcoidosis of the lung. Previous surgeries included hysteroscopy, dilation and curettage, and endoscopic lung biopsy. Patient medications included Chromagen and Claritin. For donor B, pathology indicated grade 3 adenosquamous carcinoma forming a mass within the uterine fundus and involving the anterior uterine wall, as well as focally involving an adjacent endometrial polyp. The tumor invaded to a maximum d pth of 7mm (uninvolved wall thickness, 2.2cm). The adjacent endometrium was inactive. Paraffin section immunostains for estrogen receptors and progesterone receptors were positive. Patient history included malignant breast neoplasm. Previous surgeries included unilateral extended simple mastectomy and bilateral tubal destruction. Patient medications included Megase and CAF (Cyclophosphamide, Adriamycin, Fluoroacil).

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and

d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-21.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:22-42.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

18. A method for treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional PRTS, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

29. A diagnostic test for a condition or disease associated with the expression of PRTS in a biological sample comprising the steps of:

a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of PRTS in a subject, comprising administering to said subject an effective amount of the composition of claim 31.

33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of PRTS in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

b) isolating antibodies from said animal; and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

b) isolating antibody producing cells from the animal;

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;

d) culturing the hybridoma cells; and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21 in a sample, comprising the steps of:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21 in the sample.

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21 from a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

54. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

55. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

66. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:22.

67. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:23.

68. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:24.

69. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:25.

70. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:26.

71. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:27.

72. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:28.

73. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:29.

74. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:30.

75. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:31.

76. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:32.

77. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:33.

78. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:34.

79. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:35.

80. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:36.

81. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:37.

82. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:38.

83. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:39.

84. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:40.

85. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:41.

86. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:42.

<110> INCYTE GENOMICS, INC.
YUE, Henry
ELLIOTT, Vicki
GANDHI, Ameena R.
LAL, Preeti
AU-YOUNG, Janice
TRIBOULEY, Catherine M.
DELEGEANE, Angelo M.
BAUGHN, Mariah R.
NGUYEN, Danniel B.
LEE, Ernestine A.
HAFALIA, April
KHAN, Farrah A.
WALIA, Narinder K.
YAO, Monique G.
LU, Dyung Aina M.
PATTERSON, Chandra
TANG, Y. Tom
WALSH, Roderick T.
AZIMZAI, Yalda
LU, Yan
RAMKUMAR, Jayalaximi
XU, Yuming
REDDY, Roopa
DAS, Depopriya
KEARNEY, Liam
KALLICK, Deborah A.

<120> Proteases

<130> PI-0123 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/212,336; 60/213,955; 60/215,396; 60/216,821; 60/218,946
<151> 2000-06-16; 2000-06-22; 2000-06-29; 2000-07-07; 2000-07-14

<160> 42
<170> PERL Program

<210> 1
<211> 232
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 275791CD1

<400> 1
```
Met Pro Glu Asn Pro Asp Thr Met Glu Thr Glu Lys Pro Lys Thr
  1               5                  10                  15
Ile Thr Glu Leu Asp Pro Ala Ser Phe Thr Glu Ile Thr Lys Asp
                 20                  25                  30
Cys Asp Glu Asn Lys Glu Asn Lys Thr Pro Glu Gly Ser Gln Gly
                 35                  40                  45
Glu Val Asp Trp Leu Gln Gln Tyr Asp Met Glu Arg Glu Arg Glu
                 50                  55                  60
Glu Gln Glu Leu Gln Gln Ala Leu Ala Gln Ser Leu Gln Glu Gln
                 65                  70                  75
Glu Ala Trp Glu Gln Lys Glu Asp Asp Asp Leu Lys Arg Ala Thr
                 80                  85                  90
Glu Leu Ser Leu Gln Glu Phe Asn Asn Ser Phe Val Asp Ala Leu
                 95                 100                 105
Gly Ser Asp Glu Asp Ser Gly Asn Glu Asp Val Phe Asp Met Glu
                110                 115                 120
```

```
Tyr Thr Glu Ala Glu Ala Glu Glu Leu Lys Arg Asn Ala Glu Thr
                125                 130                 135
Gly Asn Leu Pro His Ser Tyr Arg Leu Ile Ser Val Val Ser His
                140                 145                 150
Ile Gly Ser Thr Ser Ser Ser Gly His Tyr Ile Ser Asp Val Tyr
                155                 160                 165
Asp Ile Lys Lys Gln Ala Trp Phe Thr Tyr Asn Asp Leu Glu Val
                170                 175                 180
Ser Lys Ile Gln Glu Ala Ala Val Gln Ser Asp Arg Asp Arg Ser
                185                 190                 195
Gly Tyr Ile Phe Phe Tyr Met His Lys Glu Ile Phe Asp Glu Leu
                200                 205                 210
Leu Glu Thr Glu Lys Asn Ser Gln Ser Leu Ser Thr Glu Val Gly
                215                 220                 225
Lys Thr Thr Arg Gln Ala Ser
                230
```

<210> 2
<211> 365
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1389845CD1

<400> 2

```
Met Pro Lys Tyr Leu Gly Gly Gly Cys Cys Ile Pro Gly Pro Trp
  1               5                  10                  15
Ala Glu Arg Arg Val Tyr Ser Leu Gly His Gln Asp Lys Ser Arg
                 20                  25                  30
Thr His Gln Glu Leu Arg Thr Asp Arg Arg Thr Thr Glu Gly Val
                 35                  40                  45
Thr Gly Trp Cys Glu Asp Trp Cys Pro Trp Ala Arg Thr Leu Leu
                 50                  55                  60
Ser Ser Pro Cys Trp Leu Gln Thr Arg Val Gln Ala Leu Gly Ser
                 65                  70                  75
Ala Thr Leu Thr Gln Pro Ser Leu Glu Asp Arg Met Arg Gly Val
                 80                  85                  90
Ser Cys Leu Gln Val Leu Leu Leu Leu Val Leu Gly Ala Ala Gly
                 95                 100                 105
Thr Gln Gly Arg Lys Ser Ala Ala Cys Gly Gln Pro Arg Met Ser
                110                 115                 120
Ser Arg Ile Val Gly Gly Arg Asp Gly Arg Asp Gly Glu Trp Pro
                125                 130                 135
Trp Gln Ala Ser Ile Gln His Arg Gly Ala His Val Cys Gly Gly
                140                 145                 150
Ser Leu Ile Ala Pro Gln Trp Val Leu Thr Ala Ala His Cys Phe
                155                 160                 165
Pro Arg Arg Ala Leu Pro Ala Glu Tyr Arg Val Arg Leu Gly Ala
                170                 175                 180
Leu Arg Leu Gly Ser Thr Ser Pro Arg Thr Leu Ser Val Pro Val
                185                 190                 195
Arg Arg Val Leu Leu Pro Pro Asp Tyr Ser Glu Asp Gly Ala Arg
                200                 205                 210
Gly Asp Leu Ala Leu Leu Gln Leu Arg Arg Pro Val Pro Leu Ser
                215                 220                 225
Ala Arg Val Gln Pro Val Cys Leu Pro Val Pro Gly Ala Arg Pro
                230                 235                 240
Pro Pro Gly Thr Pro Cys Arg Val Thr Gly Trp Gly Ser Leu Arg
                245                 250                 255
Pro Gly Val Pro Leu Pro Glu Trp Arg Pro Leu Gln Gly Val Arg
                260                 265                 270
Val Pro Leu Leu Asp Ser Arg Thr Cys Asp Gly Leu Tyr His Val
                275                 280                 285
Gly Ala Asp Val Pro Gln Ala Glu Arg Ile Val Leu Pro Gly Ser
                290                 295                 300
```

Leu Cys Ala Gly Tyr Pro Gln Gly His Lys Asp Ala Cys Gln Val
305 310 315
Cys Thr Gln Pro Pro Gln Pro Pro Glu Ser Pro Pro Cys Ala Gln
320 325 330
His Pro Pro Ser Leu Asn Ser Arg Thr Gln Asp Ile Pro Thr Gln
335 340 345
Ala Gln Asp Pro Gly Leu Gln Pro Arg Gly Thr Thr Pro Gly Val
350 355 360
Trp Asn Pro Glu Asn
365

<210> 3
<211> 416
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1726609CD1

<400> 3
Met Trp Gly Arg Tyr Asp Ile Val Phe Leu Pro Pro Ser Phe Pro
1 5 10 15
Ile Val Ala Met Glu Asn Pro Cys Leu Thr Phe Ile Ile Ser Ser
20 25 30
Ile Leu Glu Ser Asp Glu Phe Leu Val Ile Asp Val Ile His Glu
35 40 45
Val Ala His Ser Trp Phe Gly Asn Ala Val Thr Asn Ala Thr Trp
50 55 60
Glu Glu Met Trp Leu Ser Glu Gly Leu Ala Thr Tyr Ala Gln Arg
65 70 75
Arg Ile Thr Thr Glu Thr Tyr Gly Ala Ala Phe Thr Cys Leu Glu
80 85 90
Thr Ala Phe Arg Leu Asp Ala Leu His Arg Gln Met Lys Leu Leu
95 100 105
Gly Glu Asp Ser Pro Val Ser Lys Leu Gln Val Lys Leu Glu Pro
110 115 120
Gly Val Asn Pro Ser His Leu Met Asn Leu Phe Thr Tyr Glu Lys
125 130 135
Gly Tyr Cys Phe Val Tyr Tyr Leu Ser Gln Leu Cys Gly Asp Pro
140 145 150
Gln Arg Phe Asp Asp Phe Leu Arg Ala Tyr Val Glu Lys Tyr Lys
155 160 165
Phe Thr Ser Val Val Ala Gln Asp Leu Leu Asp Ser Phe Leu Ser
170 175 180
Phe Phe Pro Glu Leu Lys Glu Gln Ser Val Asp Cys Arg Ala Gly
185 190 195
Leu Glu Phe Glu Arg Trp Leu Asn Ala Thr Gly Pro Pro Leu Ala
200 205 210
Glu Pro Asp Leu Ser Gln Gly Ser Ser Leu Thr Arg Pro Val Glu
215 220 225
Ala Leu Phe Gln Leu Trp Thr Ala Glu Pro Leu Asp Gln Ala Ala
230 235 240
Ala Ser Ala Ser Ala Ile Asp Ile Ser Lys Trp Arg Thr Phe Gln
245 250 255
Thr Ala Leu Phe Leu Asp Arg Leu Leu Asp Gly Ser Pro Leu Pro
260 265 270
Gln Glu Val Val Met Ser Leu Ser Lys Cys Tyr Ser Ser Leu Leu
275 280 285
Asp Ser Met Asn Ala Glu Ile Arg Ile Arg Trp Leu Gln Ile Val
290 295 300
Val Arg Asn Asp Tyr Tyr Pro Asp Leu His Arg Val Arg Arg Phe
305 310 315
Leu Glu Ser Gln Met Ser Arg Met Tyr Thr Ile Pro Leu Tyr Glu
320 325 330
Asp Leu Cys Thr Gly Ala Leu Lys Ser Phe Ala Leu Glu Val Phe
335 340 345

```
Tyr Gln Thr Gln Gly Arg Leu His Pro Asn Leu Arg Arg Ala Ile
                350                 355                 360
Gln Gln Ile Leu Ser Gln Gly Leu Gly Ser Ser Thr Glu Pro Ala
                365                 370                 375
Ser Glu Pro Ser Thr Glu Leu Gly Lys Ala Glu Ala Asp Thr Asp
                380                 385                 390
Ser Asp Ala Gln Ala Leu Leu Leu Gly Asp Glu Ala Pro Ser Ser
                395                 400                 405
Ala Ile Ser Leu Arg Asp Val Asn Val Ser Ala
                410                 415
```

<210> 4
<211> 714
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4503848CD1

<400> 4

```
Met His Ile His Met Leu Thr Leu Asp Gln Gln Lys Ser Leu Ile
 1               5                  10                  15
Leu Ile Leu Phe Leu Ile Leu Phe Arg Val Gly Gly Ser Arg Ile
                20                  25                  30
Leu Leu Arg Met Thr Leu Gly Arg Glu Val Met Ser Pro Leu Gln
                35                  40                  45
Ala Met Ser Ser Tyr Thr Val Ala Gly Arg Asn Val Leu Arg Trp
                50                  55                  60
Asp Leu Ser Pro Glu Gln Ile Lys Thr Arg Thr Glu Glu Leu Ile
                65                  70                  75
Val Gln Thr Lys Gln Val Tyr Asp Ala Val Gly Met Leu Gly Ile
                80                  85                  90
Glu Glu Val Thr Tyr Glu Asn Cys Leu Gln Ala Leu Ala Asp Val
                95                 100                 105
Glu Val Lys Tyr Ile Val Glu Arg Thr Met Leu Asp Phe Pro Gln
               110                 115                 120
His Val Ser Ser Asp Lys Glu Val Arg Ala Ala Ser Thr Glu Ala
               125                 130                 135
Asp Lys Arg Leu Ser Arg Phe Asp Ile Glu Met Ser Met Arg Gly
               140                 145                 150
Asp Ile Phe Glu Arg Ile Val His Leu Gln Glu Thr Cys Asp Leu
               155                 160                 165
Gly Lys Ile Lys Pro Glu Ala Arg Arg Tyr Leu Glu Lys Ser Ile
               170                 175                 180
Lys Met Gly Lys Arg Asn Gly Leu His Leu Pro Glu Gln Val Gln
               185                 190                 195
Asn Glu Ile Lys Ser Met Lys Lys Arg Met Ser Glu Leu Cys Ile
               200                 205                 210
Asp Phe Asn Lys Asn Leu Asn Glu Asp Asp Thr Phe Leu Val Phe
               215                 220                 225
Ser Lys Ala Glu Leu Gly Ala Leu Pro Asp Asp Phe Ile Asp Ser
               230                 235                 240
Leu Glu Lys Thr Asp Asp Asp Lys Tyr Lys Ile Thr Leu Lys Tyr
               245                 250                 255
Pro His Tyr Phe Pro Val Met Lys Lys Cys Cys Ile Pro Glu Thr
               260                 265                 270
Arg Arg Arg Met Glu Met Ala Phe Asn Thr Arg Cys Lys Glu Glu
               275                 280                 285
Asn Thr Ile Ile Leu Gln Gln Leu Leu Pro Leu Arg Thr Lys Val
               290                 295                 300
Ala Lys Leu Leu Gly Tyr Ser Thr His Ala Asp Phe Val Leu Glu
               305                 310                 315
Met Asn Thr Ala Lys Ser Thr Ser Arg Val Thr Ala Phe Leu Asp
               320                 325                 330
Asp Leu Ser Gln Lys Leu Lys Pro Leu Gly Glu Ala Glu Arg Glu
               335                 340                 345
```

```
Phe Ile Leu Asn Leu Lys Lys Lys Glu Cys Lys Asp Arg Gly Phe
                350                 355                 360
Glu Tyr Asp Gly Lys Ile Asn Ala Trp Asp Leu Tyr Tyr Tyr Met
                365                 370                 375
Thr Gln Thr Glu Glu Leu Lys Tyr Ser Ile Asp Gln Glu Phe Leu
                380                 385                 390
Lys Glu Tyr Phe Pro Ile Glu Val Val Thr Glu Gly Leu Leu Asn
                395                 400                 405
Thr Tyr Gln Glu Leu Leu Gly Leu Ser Phe Glu Gln Met Thr Asp
                410                 415                 420
Ala His Val Trp Asn Lys Ser Val Thr Leu Tyr Thr Val Lys Asp
                425                 430                 435
Lys Ala Thr Gly Glu Val Leu Gly Gln Phe Tyr Leu Asp Leu Tyr
                440                 445                 450
Pro Arg Glu Gly Lys Tyr Asn His Ala Ala Cys Phe Gly Leu Gln
                455                 460                 465
Pro Gly Cys Leu Leu Pro Asp Gly Ser Arg Met Met Ala Val Ala
                470                 475                 480
Ala Leu Val Val Asn Phe Ser Gln Pro Val Ala Gly Arg Pro Ser
                485                 490                 495
Leu Leu Arg His Asp Glu Val Arg Thr Tyr Phe His Glu Phe Gly
                500                 505                 510
His Val Met His Gln Ile Cys Ala Gln Thr Asp Phe Ala Arg Phe
                515                 520                 525
Ser Gly Thr Asn Val Glu Thr Asp Phe Val Glu Val Pro Ser Gln
                530                 535                 540
Met Leu Glu Asn Trp Val Trp Asp Val Asp Ser Leu Arg Arg Leu
                545                 550                 555
Ser Lys His Tyr Lys Asp Gly Ser Pro Ile Ala Asp Asp Leu Leu
                560                 565                 570
Glu Lys Leu Val Ala Ser Arg Leu Val Asn Thr Gly Leu Leu Thr
                575                 580                 585
Leu Arg Gln Ile Val Leu Ser Lys Val Asp Gln Ser Leu His Thr
                590                 595                 600
Asn Thr Ser Leu Asp Ala Ala Ser Glu Tyr Ala Lys Tyr Cys Ser
                605                 610                 615
Glu Ile Leu Gly Val Ala Ala Thr Pro Gly Thr Asn Met Pro Ala
                620                 625                 630
Thr Phe Gly His Leu Ala Gly Gly Tyr Asp Gly Gln Tyr Tyr Gly
                635                 640                 645
Tyr Leu Trp Ser Glu Val Phe Ser Met Asp Met Phe Tyr Ser Cys
                650                 655                 660
Phe Lys Lys Glu Gly Ile Met Asn Pro Glu Val Gly Met Lys Tyr
                665                 670                 675
Arg Asn Leu Ile Leu Lys Pro Gly Gly Ser Leu Asp Gly Met Asp
                680                 685                 690
Met Leu His Asn Phe Leu Lys Arg Glu Pro Asn Gln Lys Ala Phe
                695                 700                 705
Leu Met Ser Arg Gly Leu His Ala Pro
                710
```

```
<210> 5
<211> 367
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5544089CD1

<400> 5
Met Phe Ala Pro Ser Val Leu Ser Ser Gly Leu Ser Gly Gly Ala
 1               5                  10                  15
Ser Lys Gly Arg Lys Met Glu Leu Ile Gln Pro Lys Glu Pro Thr
                20                  25                  30
Ser Gln Tyr Ile Ser Leu Cys His Glu Leu His Thr Leu Phe Gln
                35                  40                  45
```

```
Val Met Trp Ser Gly Lys Trp Ala Leu Val Ser Pro Phe Ala Met
                50                  55                  60
Leu His Ser Val Trp Arg Leu Ile Pro Ala Phe Arg Gly Tyr Ala
                65                  70                  75
Gln Gln Asp Ala Gln Glu Phe Leu Cys Glu Leu Leu Asp Lys Ile
                80                  85                  90
Gln Arg Glu Leu Glu Thr Thr Gly Thr Ser Leu Pro Ala Leu Ile
                95                  100                 105
Pro Thr Ser Gln Arg Lys Leu Ile Lys Gln Val Leu Asn Val Val
                110                 115                 120
Asn Asn Ile Phe His Gly Gln Leu Leu Ser Gln Val Thr Cys Leu
                125                 130                 135
Ala Cys Asp Asn Lys Ser Asn Thr Ile Glu Pro Phe Trp Asp Leu
                140                 145                 150
Ser Leu Glu Phe Pro Glu Arg Tyr Gln Cys Ser Gly Lys Asp Ile
                155                 160                 165
Ala Ser Gln Pro Cys Leu Val Thr Glu Met Leu Ala Lys Phe Thr
                170                 175                 180
Glu Thr Glu Ala Leu Glu Gly Lys Ile Tyr Val Cys Asp Gln Cys
                185                 190                 195
Asn Ser Lys Arg Arg Arg Phe Ser Ser Lys Pro Val Val Leu Thr
                200                 205                 210
Glu Ala Gln Lys Gln Leu Met Ile Cys His Leu Pro Gln Val Leu
                215                 220                 225
Arg Leu His Leu Lys Arg Phe Arg Trp Ser Gly Arg Asn Asn Arg
                230                 235                 240
Glu Lys Ile Gly Val His Val Gly Phe Glu Glu Ile Leu Asn Met
                245                 250                 255
Glu Pro Tyr Cys Cys Arg Glu Thr Leu Lys Ser Leu Arg Pro Glu
                260                 265                 270
Cys Phe Ile Tyr Asp Leu Ser Ala Val Val Met His His Gly Lys
                275                 280                 285
Gly Phe Gly Ser Gly His Tyr Thr Ala Tyr Cys Tyr Asn Ser Glu
                290                 295                 300
Gly Gly Phe Trp Val His Cys Asn Asp Ser Lys Leu Ser Met Cys
                305                 310                 315
Thr Met Asp Glu Val Cys Lys Ala Gln Ala Tyr Ile Leu Phe Tyr
                320                 325                 330
Thr Gln Arg Val Thr Glu Asn Gly His Ser Lys Leu Leu Pro Pro
                335                 340                 345
Glu Leu Leu Leu Gly Ser Gln His Pro Asn Glu Asp Ala Asp Thr
                350                 355                 360
Ser Ser Asn Glu Ile Leu Ser
                365
```

<210> 6
<211> 235
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474081CD1

<400> 6

```
Met Lys Tyr Val Phe Tyr Leu Gly Val Leu Ala Gly Thr Phe Phe
 1              5                   10                  15
Phe Ala Asp Ser Ser Val Gln Lys Glu Asp Pro Ala Pro Tyr Leu
                20                  25                  30
Val Tyr Leu Lys Ser His Phe Asn Pro Cys Val Gly Val Leu Ile
                35                  40                  45
Lys Pro Ser Trp Val Leu Ala Pro Ala His Cys Tyr Leu Pro Asn
                50                  55                  60
Leu Lys Val Met Leu Gly Asn Phe Lys Ser Arg Val Arg Asp Gly
                65                  70                  75
Thr Glu Gln Thr Ile Asn Pro Ile Gln Ile Val Arg Tyr Trp Asn
                80                  85                  90
```

```
Tyr Ser His Ser Ala Pro Gln Asp Asp Leu Met Leu Ile Lys Leu
                95                  100                 105
Ala Lys Pro Ala Met Leu Asn Pro Lys Val Gln Pro Leu Thr Leu
                110                 115                 120
Ala Thr Thr Asn Val Arg Pro Gly Thr Val Cys Leu Leu Ser Gly
                125                 130                 135
Leu Asp Trp Ser Gln Glu Asn Ser Gly Arg His Pro Asp Leu Arg
                140                 145                 150
Gln Asn Leu Glu Ala Pro Val Met Ser Asp Arg Glu Cys Gln Lys
                155                 160                 165
Thr Glu Gln Gly Lys Ser His Arg Asn Ser Leu Cys Val Lys Phe
                170                 175                 180
Val Lys Val Phe Ser Arg Ile Phe Gly Glu Val Ala Val Ala Thr
                185                 190                 195
Val Ile Cys Lys Asp Lys Leu Gln Gly Ile Glu Val Gly His Phe
                200                 205                 210
Met Gly Gly Asp Val Gly Ile Tyr Thr Asn Val Tyr Lys Tyr Val
                215                 220                 225
Ser Trp Ile Glu Asn Thr Ala Lys Asp Lys
                230                 235
```

<210> 7
<211> 488
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5281209CD1

<400> 7

```
Met Gln Pro Thr Gly Arg Glu Gly Ser Arg Ala Leu Ser Arg Arg
 1              5                   10                  15
Tyr Leu Arg Arg Leu Leu Leu Leu Leu Leu Leu Leu Leu Leu Arg
                20                  25                  30
Gln Pro Val Thr Arg Ala Glu Thr Thr Pro Gly Ala Pro Arg Ala
                35                  40                  45
Leu Ser Thr Leu Gly Ser Pro Ser Leu Phe Thr Thr Pro Gly Val
                50                  55                  60
Pro Ser Ala Leu Thr Thr Pro Gly Leu Thr Thr Pro Gly Thr Pro
                65                  70                  75
Lys Thr Leu Asp Leu Arg Gly Arg Ala Gln Ala Leu Met Arg Ser
                80                  85                  90
Phe Pro Leu Val Asp Gly His Asn Asp Leu Pro Gln Val Leu Arg
                95                  100                 105
Gln Arg Tyr Lys Asn Val Leu Gln Asp Val Asn Leu Arg Asn Phe
                110                 115                 120
Ser His Gly Gln Thr Ser Leu Asp Arg Leu Arg Asp Gly Leu Val
                125                 130                 135
Gly Ala Gln Phe Trp Ser Ala Ser Val Ser Cys Gln Ser Gln Asp
                140                 145                 150
Gln Thr Ala Val Arg Leu Ala Leu Glu Gln Ile Asp Leu Ile His
                155                 160                 165
Arg Met Cys Ala Ser Tyr Ser Glu Leu Glu Leu Val Thr Ser Ala
                170                 175                 180
Glu Gly Leu Asn Ser Ser Gln Lys Leu Ala Cys Leu Ile Gly Val
                185                 190                 195
Glu Gly Gly His Ser Leu Asp Ser Ser Leu Ser Val Leu Arg Ser
                200                 205                 210
Phe Tyr Val Leu Gly Val Arg Tyr Leu Thr Leu Thr Phe Thr Cys
                215                 220                 225
Ser Thr Pro Trp Ala Glu Ser Ser Thr Lys Phe Arg His His Met
                230                 235                 240
Tyr Thr Asn Val Ser Gly Leu Thr Ser Phe Gly Glu Lys Val Val
                245                 250                 255
Glu Glu Leu Asn Arg Leu Gly Met Met Ile Asp Leu Ser Tyr Ala
                260                 265                 270
```

```
Ser Asp Thr Leu Ile Arg Arg Val Leu Glu Val Ser Gln Ala Pro
                275                 280                 285
Val Ile Phe Ser His Ser Ala Ala Arg Ala Val Cys Asp Asn Leu
                290                 295                 300
Leu Asn Val Pro Asp Asp Ile Leu Gln Leu Leu Lys Lys Asn Gly
                305                 310                 315
Gly Ile Val Met Val Thr Leu Ser Met Gly Val Leu Gln Cys Asn
                320                 325                 330
Leu Leu Ala Asn Val Ser Thr Val Ala Asp His Phe Asp His Ile
                335                 340                 345
Arg Ala Val Ile Gly Ser Glu Phe Ile Gly Ile Gly Gly Asn Tyr
                350                 355                 360
Asp Gly Thr Gly Arg Phe Pro Gln Gly Leu Glu Asp Val Ser Thr
                365                 370                 375
Tyr Pro Val Leu Ile Glu Glu Leu Leu Ser Arg Ser Trp Ser Glu
                380                 385                 390
Glu Glu Leu Gln Gly Val Leu Arg Gly Asn Leu Leu Arg Val Phe
                395                 400                 405
Arg Gln Val Glu Lys Val Arg Glu Glu Ser Arg Ala Gln Ser Pro
                410                 415                 420
Val Glu Ala Glu Phe Pro Tyr Gly Gln Leu Ser Thr Ser Cys His
                425                 430                 435
Ser His Leu Val Pro Gln Asn Gly His Gln Ala Thr His Leu Glu
                440                 445                 450
Val Thr Lys Gln Pro Thr Asn Arg Val Pro Trp Arg Ser Ser Asn
                455                 460                 465
Ala Ser Pro Tyr Leu Val Pro Gly Leu Val Ala Ala Ala Thr Ile
                470                 475                 480
Pro Thr Phe Thr Gln Trp Leu Cys
                485
```

<210> 8
<211> 346
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2256251CD1

<400> 8

```
Met Leu Leu Gly Arg Val Trp Gln Thr Arg Glu Leu Lys Ser Lys
 1               5                  10                  15
Val Pro Lys Lys Ala Gly Arg Cys Gly Gln Gly Arg Leu His Gly
                20                  25                  30
Gly Ser Ala Val Gly Phe Leu Gly Ser Pro Pro Gly Thr Pro Ser
                35                  40                  45
Ser Phe Asp Leu Gly Cys Gly Arg Pro Gln Val Ser Asp Ala Gly
                50                  55                  60
Gly Arg Ile Val Gly Gly His Ala Ala Pro Ala Gly Ala Trp Pro
                65                  70                  75
Trp Gln Ala Ser Leu Arg Leu Arg Arg Val His Val Cys Gly Gly
                80                  85                  90
Ser Leu Leu Ser Pro Gln Trp Val Leu Thr Ala Ala His Cys Phe
                95                 100                 105
Ser Gly Ser Leu Asn Ser Ser Asp Tyr Gln Val His Leu Gly Glu
                110                 115                 120
Leu Glu Ile Thr Leu Ser Pro His Phe Ser Thr Val Arg Gln Ile
                125                 130                 135
Ile Leu His Ser Ser Pro Ser Gly Gln Pro Gly Thr Ser Gly Asp
                140                 145                 150
Ile Ala Leu Val Glu Leu Ser Val Pro Val Thr Leu Phe Ser Arg
                155                 160                 165
Ile Leu Pro Val Cys Leu Pro Glu Ala Ser Asp Asp Phe Cys Pro
                170                 175                 180
Gly Ile Arg Cys Trp Val Thr Gly Trp Gly Tyr Thr Arg Glu Gly
                185                 190                 195
```

Glu Pro Leu Pro Pro Pro Tyr Ser Leu Arg Glu Val Lys Val Ser
200 205 210
Val Val Asp Thr Glu Thr Cys Arg Arg Asp Tyr Pro Gly Pro Gly
215 220 225
Gly Ser Ile Leu Gln Pro Asp Met Leu Cys Ala Arg Gly Pro Gly
230 235 240
Asp Ala Cys Gln Asp Asp Ser Gly Gly Pro Leu Val Cys Gln Val
245 250 255
Asn Gly Ala Trp Val Gln Ala Gly Ile Val Ser Trp Gly Glu Gly
260 265 270
Cys Gly Arg Pro Asn Arg Pro Gly Val Tyr Thr Arg Val Pro Ala
275 280 285
Tyr Val Asn Trp Ile Arg Arg His Ile Thr Ala Ser Gly Gly Ser
290 295 300
Glu Ser Gly Tyr Pro Arg Leu Pro Leu Leu Ala Gly Leu Phe Leu
305 310 315
Pro Gly Leu Phe Leu Leu Leu Val Ser Cys Val Leu Leu Ala Lys
320 325 330
Cys Leu Leu His Pro Ser Ala Asp Gly Thr Pro Phe Pro Ala Pro
335 340 345
Asp

<210> 9
<211> 882
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7160544CD1

<400> 9
Met Ala Ala Ala Met Glu Thr Glu Gln Leu Gly Val Glu Ile Phe
1 5 10 15
Glu Thr Ala Asp Cys Glu Glu Asn Ile Glu Ser Gln Asp Arg Pro
20 25 30
Lys Leu Glu Pro Phe Tyr Val Glu Arg Tyr Ser Trp Ser Gln Leu
35 40 45
Lys Lys Leu Leu Ala Asp Thr Arg Lys Tyr His Gly Tyr Met Met
50 55 60
Ala Lys Ala Pro His Asp Phe Met Phe Val Lys Arg Asn Asp Pro
65 70 75
Asp Gly Pro His Ser Asp Arg Ile Tyr Tyr Leu Ala Met Ser Gly
80 85 90
Glu Asn Arg Glu Asn Thr Leu Phe Tyr Ser Glu Ile Pro Lys Thr
95 100 105
Ile Asn Arg Ala Ala Val Leu Met Leu Ser Trp Lys Pro Leu Leu
110 115 120
Asp Leu Phe Gln Ala Thr Leu Asp Tyr Gly Met Tyr Ser Arg Glu
125 130 135
Glu Glu Leu Leu Arg Glu Arg Lys Arg Ile Gly Thr Val Gly Ile
140 145 150
Ala Ser Tyr Asp Tyr His Gln Gly Ser Gly Thr Phe Leu Phe Gln
155 160 165
Ala Gly Ser Gly Ile Tyr His Val Lys Asp Gly Gly Pro Gln Gly
170 175 180
Phe Thr Gln Gln Pro Leu Arg Pro Asn Leu Val Glu Thr Ser Cys
185 190 195
Pro Asn Ile Arg Met Asp Pro Lys Leu Cys Pro Ala Asp Pro Asp
200 205 210
Trp Ile Ala Phe Ile His Ser Asn Asp Ile Trp Ile Ser Asn Ile
215 220 225
Val Thr Arg Glu Glu Arg Arg Leu Thr Tyr Val His Asn Glu Leu
230 235 240
Ala Asn Met Glu Glu Asp Ala Arg Ser Ala Gly Val Ala Thr Phe
245 250 255

Val Leu Gln Glu Glu Phe Asp Arg Tyr Ser Gly Tyr Trp Trp Cys
260 265 270
Pro Lys Ala Glu Thr Thr Pro Ser Gly Gly Lys Ile Leu Arg Ile
275 280 285
Leu Tyr Glu Glu Asn Asp Glu Ser Glu Val Glu Ile Ile His Val
290 295 300
Thr Ser Pro Met Leu Glu Thr Arg Arg Ala Asp Ser Phe Arg Tyr
305 310 315
Pro Lys Thr Gly Thr Ala Asn Pro Lys Val Thr Phe Lys Met Ser
320 325 330
Glu Ile Met Ile Asp Ala Glu Gly Arg Ile Ile Asp Val Ile Asp
335 340 345
Lys Glu Leu Ile Gln Pro Phe Glu Ile Leu Phe Glu Gly Val Glu
350 355 360
Tyr Ile Ala Arg Ala Gly Trp Thr Pro Glu Gly Lys Tyr Ala Trp
365 370 375
Ser Ile Leu Leu Asp Arg Ser Gln Thr Arg Leu Gln Ile Val Leu
380 385 390
Ile Ser Pro Glu Leu Phe Ile Pro Val Glu Asp Asp Val Met Glu
395 400 405
Arg Gln Arg Leu Ile Glu Ser Val Pro Asp Ser Val Thr Pro Leu
410 415 420
Ile Ile Tyr Glu Glu Thr Thr Asp Ile Trp Ile Asn Ile His Asp
425 430 435
Ile Phe His Val Phe Pro Gln Ser His Glu Glu Glu Ile Glu Phe
440 445 450
Ile Phe Ala Ser Glu Cys Lys Thr Gly Phe Arg His Leu Tyr Lys
455 460 465
Ile Thr Ser Ile Leu Lys Glu Ser Lys Tyr Lys Arg Ser Ser Gly
470 475 480
Gly Leu Pro Ala Pro Ser Asp Phe Lys Cys Pro Ile Lys Glu Glu
485 490 495
Ile Ala Ile Thr Ser Gly Glu Trp Glu Val Leu Gly Arg His Gly
500 505 510
Ser Asn Ile Gln Val Asp Glu Val Arg Arg Leu Val Tyr Phe Glu
515 520 525
Gly Thr Lys Asp Ser Pro Leu Glu His His Leu Tyr Val Val Ser
530 535 540
Tyr Val Asn Pro Gly Glu Val Thr Arg Leu Thr Asp Arg Gly Tyr
545 550 555
Ser His Ser Cys Cys Ile Ser Gln His Cys Asp Phe Phe Ile Ser
560 565 570
Lys Tyr Ser Asn Gln Lys Asn Pro His Cys Val Ser Leu Tyr Lys
575 580 585
Leu Ser Ser Pro Glu Asp Asp Pro Thr Cys Lys Thr Lys Glu Phe
590 595 600
Trp Ala Thr Ile Leu Asp Ser Ala Gly Pro Leu Pro Asp Tyr Thr
605 610 615
Pro Pro Glu Ile Phe Ser Phe Glu Ser Thr Thr Gly Phe Thr Leu
620 625 630
Tyr Gly Met Leu Tyr Lys Pro His Asp Leu Gln Pro Gly Lys Lys
635 640 645
Tyr Pro Thr Val Leu Phe Ile Tyr Gly Gly Pro Gln Val Gln Leu
650 655 660
Val Asn Asn Arg Phe Lys Gly Val Lys Tyr Phe Arg Leu Asn Thr
665 670 675
Leu Ala Ser Leu Gly Tyr Val Val Val Val Ile Asp Asn Arg Gly
680 685 690
Ser Cys His Arg Gly Leu Lys Phe Glu Gly Ala Phe Lys Tyr Lys
695 700 705
Met Gly Gln Ile Glu Ile Asp Asp Gln Val Glu Gly Leu Gln Tyr
710 715 720
Leu Ala Ser Arg Tyr Asp Phe Ile Asp Leu Asp Arg Val Gly Ile
725 730 735
His Gly Trp Ser Tyr Gly Gly Tyr Leu Ser Leu Met Ala Leu Met
740 745 750
Gln Arg Ser Asp Ile Phe Arg Val Ala Ile Ala Gly Ala Pro Val

755 760 765
Thr Leu Trp Ile Phe Tyr Asp Thr Gly Tyr Thr Glu Arg Tyr Met
770 775 780
Gly His Pro Asp Gln Asn Glu Gln Gly Tyr Tyr Leu Gly Ser Val
785 790 795
Ala Met Gln Ala Glu Lys Phe Pro Ser Glu Pro Asn Arg Leu Leu
800 805 810
Leu Leu His Gly Phe Leu Asp Glu Asn Val His Phe Ala His Thr
815 820 825
Ser Ile Leu Leu Ser Phe Leu Val Arg Ala Gly Lys Pro Tyr Asp
830 835 840
Leu Gln Ile Tyr Pro Gln Glu Arg His Ser Ile Arg Val Pro Glu
845 850 855
Ser Gly Glu His Tyr Glu Leu His Leu Leu His Tyr Leu Gln Glu
860 865 870
Asn Leu Gly Ser Arg Ile Ala Ala Leu Lys Val Ile
875 880

<210> 10
<211> 1189
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477386CD1

<400> 10
Met Ala Pro Leu Arg Ala Leu Leu Ser Tyr Leu Leu Pro Leu His
1 5 10 15
Cys Ala Leu Cys Ala Ala Ala Gly Ser Arg Thr Pro Glu Leu His
20 25 30
Leu Ser Gly Lys Leu Ser Asp Tyr Gly Val Thr Val Pro Cys Ser
35 40 45
Thr Asp Phe Arg Gly Arg Phe Leu Ser His Val Val Ser Gly Pro
50 55 60
Ala Ala Ala Ser Ala Gly Ser Met Val Val Asp Thr Pro Pro Thr
65 70 75
Leu Pro Arg His Ser Ser His Leu Arg Val Ala Arg Ser Pro Leu
80 85 90
His Pro Gly Gly Thr Leu Trp Pro Gly Arg Val Gly Arg His Ser
95 100 105
Leu Tyr Phe Asn Val Thr Val Phe Gly Lys Glu Leu His Leu Arg
110 115 120
Leu Arg Pro Asn Arg Arg Leu Val Val Pro Gly Ser Ser Val Glu
125 130 135
Trp Gln Glu Asp Phe Arg Glu Leu Phe Arg Gln Pro Leu Arg Gln
140 145 150
Glu Cys Val Tyr Thr Gly Gly Val Thr Gly Met Pro Gly Ala Ala
155 160 165
Val Ala Ile Ser Asn Cys Asp Gly Leu Ala Gly Leu Ile Arg Thr
170 175 180
Asp Ser Thr Asp Phe Phe Ile Glu Pro Leu Glu Arg Gly Gln Gln
185 190 195
Glu Lys Glu Ala Ser Gly Arg Thr His Val Val Tyr Arg Arg Glu
200 205 210
Ala Val Gln Gln Glu Trp Ala Glu Pro Asp Gly Asp Leu His Asn
215 220 225
Glu Ala Phe Gly Leu Gly Asp Leu Pro Asn Leu Leu Gly Leu Val
230 235 240
Gly Asp Gln Leu Gly Asp Thr Glu Arg Lys Arg Arg His Ala Lys
245 250 255
Pro Gly Ser Tyr Ser Ile Glu Val Leu Leu Val Val Asp Asp Ser
260 265 270
Val Val Arg Phe His Gly Lys Glu His Val Gln Asn Tyr Val Leu
275 280 285
Thr Leu Met Asn Ile Val Val Asp Glu Ile Tyr His Asp Glu Ser

```
                290                 295                 300
Leu Gly Val His Ile Asn Ile Ala Leu Val Arg Leu Ile Met Val
                305                 310                 315
Gly Tyr Arg Gln Gln Ser Leu Ser Leu Ile Glu Arg Gly Asn Pro
                320                 325                 330
Ser Arg Ser Leu Glu Gln Val Cys Arg Trp Ala His Ser Gln Gln
                335                 340                 345
Arg Gln Asp Pro Ser His Ala Glu His His Asp His Val Val Phe
                350                 355                 360
Leu Thr Arg Gln Asp Phe Gly Pro Ser Gly Gly Tyr Ala Pro Val
                365                 370                 375
Thr Gly Met Cys His Pro Leu Arg Ser Cys Ala Leu Asn His Glu
                380                 385                 390
Asp Gly Phe Ser Ser Ala Phe Val Ile Ala His Glu Thr Gly His
                395                 400                 405
Val Leu Gly Met Glu His Asp Gly Gln Gly Asn Gly Cys Ala Asp
                410                 415                 420
Glu Thr Ser Leu Gly Ser Val Met Ala Pro Leu Val Gln Ala Ala
                425                 430                 435
Phe His Arg Phe His Trp Ser Arg Cys Ser Lys Leu Glu Leu Ser
                440                 445                 450
Arg Tyr Leu Pro Ser Tyr Asp Cys Leu Leu Asp Asp Pro Phe Asp
                455                 460                 465
Pro Ala Trp Pro Gln Pro Pro Glu Leu Pro Gly Ile Asn Tyr Ser
                470                 475                 480
Met Asp Glu Gln Cys Arg Phe Asp Phe Gly Ser Gly Tyr Gln Thr
                485                 490                 495
Cys Leu Ala Phe Arg Thr Phe Glu Pro Cys Lys Gln Leu Trp Cys
                500                 505                 510
Ser His Pro Asp Asn Pro Tyr Phe Cys Lys Thr Lys Lys Gly Pro
                515                 520                 525
Pro Leu Asp Gly Thr Glu Cys Ala Pro Gly Lys Trp Cys Phe Lys
                530                 535                 540
Gly His Cys Ile Trp Lys Ser Pro Glu Gln Thr Tyr Gly Gln Asp
                545                 550                 555
Gly Gly Trp Ser Ser Trp Thr Lys Phe Gly Ser Cys Ser Arg Ser
                560                 565                 570
Cys Gly Gly Gly Val Arg Ser Arg Ser Arg Ser Cys Asn Asn Pro
                575                 580                 585
Ser Pro Ala Tyr Gly Gly Arg Leu Cys Leu Gly Pro Met Phe Glu
                590                 595                 600
Tyr Gln Val Cys Asn Ser Glu Glu Cys Pro Gly Thr Tyr Glu Asp
                605                 610                 615
Phe Arg Ala Gln Gln Cys Ala Lys Arg Asn Ser Tyr Tyr Val His
                620                 625                 630
Gln Asn Ala Lys His Ser Trp Val Pro Tyr Glu Pro Asp Asp Asp
                635                 640                 645
Ala Gln Lys Cys Glu Leu Ile Cys Gln Ser Ala Asp Thr Gly Asp
                650                 655                 660
Val Val Phe Met Asn Gln Val Val His Asp Gly Thr Arg Cys Ser
                665                 670                 675
Tyr Arg Asp Pro Tyr Ser Val Cys Ala Arg Gly Glu Cys Val Pro
                680                 685                 690
Val Gly Cys Asp Lys Glu Val Gly Ser Met Lys Ala Asp Asp Lys
                695                 700                 705
Cys Gly Val Cys Gly Gly Asp Asn Ser His Cys Arg Thr Val Lys
                710                 715                 720
Gly Thr Leu Gly Lys Ala Ser Lys Gln Ala Gly Ala Leu Lys Leu
                725                 730                 735
Val Gln Ile Pro Ala Gly Ala Arg His Ile Gln Ile Glu Ala Leu
                740                 745                 750
Glu Lys Ser Pro His Arg Ile Val Val Lys Asn Gln Val Thr Gly
                755                 760                 765
Ser Phe Ile Leu Asn Pro Lys Gly Lys Glu Ala Thr Ser Arg Thr
                770                 775                 780
Phe Thr Ala Met Gly Leu Glu Trp Glu Asp Ala Val Glu Asp Ala
                785                 790                 795
```

Lys Glu Ser Leu Lys Thr Ser Gly Pro Leu Pro Glu Ala Ile Ala
800 805 810
Ile Leu Ala Leu Pro Pro Thr Glu Gly Gly Pro Arg Ser Ser Leu
815 820 825
Ala Tyr Lys Tyr Val Ile His Glu Asp Leu Leu Pro Leu Ile Gly
830 835 840
Ser Asn Asn Val Leu Leu Glu Glu Met Asp Thr Tyr Glu Trp Ala
845 850 855
Leu Lys Ser Trp Ala Pro Cys Ser Lys Ala Cys Gly Gly Gly Ile
860 865 870
Gln Phe Thr Lys Tyr Gly Cys Arg Arg Arg Arg Asp His His Met
875 880 885
Val Gln Arg His Leu Cys Asp His Lys Lys Arg Pro Lys Pro Ile
890 895 900
Arg Arg Arg Cys Asn Gln His Pro Cys Ser Gln Pro Val Trp Val
905 910 915
Thr Glu Glu Trp Gly Ala Cys Ser Arg Ser Cys Gly Lys Leu Gly
920 925 930
Val Gln Thr Arg Gly Ile Gln Cys Leu Leu Pro Leu Ser Asn Gly
935 940 945
Thr His Lys Val Met Pro Ala Lys Ala Cys Ala Gly Asp Arg Pro
950 955 960
Glu Ala Arg Arg Pro Cys Leu Arg Val Pro Cys Pro Ala Gln Trp
965 970 975
Arg Leu Gly Ala Trp Ser Gln Cys Ser Ala Thr Cys Gly Glu Gly
980 985 990
Ile Gln Gln Arg Gln Val Val Cys Arg Thr Asn Ala Asn Ser Leu
995 1000 1005
Gly His Cys Glu Gly Asp Arg Pro Asp Thr Val Gln Val Cys Ser
1010 1015 1020
Leu Pro Ala Cys Gly Ala Glu Pro Cys Thr Gly Asp Arg Ser Val
1025 1030 1035
Phe Cys Gln Met Glu Val Leu Asp Arg Tyr Cys Ser Ile Pro Gly
1040 1045 1050
Tyr His Arg Leu Cys Cys Val Ser Cys Ile Lys Lys Ala Ser Gly
1055 1060 1065
Pro Asn Pro Gly Pro Asp Pro Gly Pro Thr Ser Leu Pro Pro Phe
1070 1075 1080
Ser Thr Pro Gly Ser Pro Leu Pro Gly Pro Gln Asp Pro Ala Asp
1085 1090 1095
Ala Ala Glu Pro Pro Gly Lys Pro Thr Gly Ser Glu Asp His Gln
1100 1105 1110
His Gly Arg Ala Thr Gln Leu Pro Gly Ala Leu Asp Thr Ser Ser
1115 1120 1125
Pro Gly Thr Gln His Pro Phe Ala Pro Glu Thr Pro Ile Pro Gly
1130 1135 1140
Ala Ser Trp Ser Ile Ser Pro Thr Thr Pro Gly Gly Leu Pro Trp
1145 1150 1155
Gly Trp Thr Gln Thr Pro Thr Pro Val Pro Glu Asp Lys Gly Gln
1160 1165 1170
Pro Gly Glu Asp Leu Arg His Pro Gly Thr Ser Leu Pro Ala Ala
1175 1180 1185
Ser Pro Val Thr

<210> 11
<211> 952
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473089CD1

<400> 11
Met Leu Leu Leu Gly Ile Leu Thr Leu Ala Phe Ala Gly Arg Thr
1 5 10 15

Ala Gly Gly Ser Glu Pro Glu Arg Glu Val Val Val Pro Ile Arg
20 25 30
Leu Asp Pro Asp Ile Asn Gly Arg Arg Tyr Tyr Trp Arg Gly Pro
35 40 45
Glu Asp Ser Gly Asp Gln Gly Leu Ile Phe Gln Ile Thr Ala Phe
50 55 60
Gln Glu Asp Phe Tyr Leu His Leu Thr Pro Asp Ala Gln Phe Leu
65 70 75
Ala Pro Ala Phe Ser Thr Glu His Leu Gly Val Pro Leu Gln Gly
80 85 90
Leu Thr Gly Gly Ser Ser Asp Leu Arg Arg Cys Phe Tyr Ser Gly
95 100 105
Asp Val Asn Ala Glu Pro Asp Ser Phe Ala Ala Val Ser Leu Cys
110 115 120
Gly Gly Leu Arg Gly Ala Phe Gly Tyr Arg Gly Ala Glu Tyr Val
125 130 135
Ile Ser Pro Leu Pro Asn Ala Ser Ala Pro Ala Ala Gln Arg Asn
140 145 150
Ser Gln Gly Ala His Leu Leu Gln Arg Arg Gly Val Pro Gly Gly
155 160 165
Pro Ser Gly Asp Pro Thr Ser Arg Cys Gly Val Ala Ser Gly Trp
170 175 180
Asn Pro Ala Ile Leu Arg Ala Leu Asp Pro Tyr Lys Pro Arg Arg
185 190 195
Ala Gly Phe Gly Glu Ser Arg Ser Arg Arg Arg Ser Gly Arg Ala
200 205 210
Lys Arg Phe Val Ser Ile Pro Arg Tyr Val Glu Thr Leu Val Val
215 220 225
Ala Asp Glu Ser Met Val Lys Phe His Gly Ala Asp Leu Glu His
230 235 240
Tyr Leu Leu Thr Leu Leu Ala Thr Ala Ala Arg Leu Tyr Arg His
245 250 255
Pro Ser Ile Leu Asn Pro Ile Asn Ile Val Val Val Lys Val Leu
260 265 270
Leu Leu Arg Asp Arg Asp Ser Gly Pro Lys Val Thr Gly Asn Ala
275 280 285
Ala Leu Thr Leu Arg Asn Phe Cys Ala Trp Gln Lys Lys Leu Asn
290 295 300
Lys Val Ser Asp Lys His Pro Glu Tyr Trp Asp Thr Ala Ile Leu
305 310 315
Phe Thr Arg Gln Asp Leu Cys Gly Ala Thr Thr Cys Asp Thr Leu
320 325 330
Gly Met Ala Asp Val Gly Thr Met Cys Asp Pro Lys Arg Ser Cys
335 340 345
Ser Val Ile Glu Asp Asp Gly Leu Pro Ser Ala Phe Thr Thr Ala
350 355 360
His Glu Leu Gly His Val Phe Asn Met Pro His Asp Asn Val Lys
365 370 375
Val Cys Glu Glu Val Phe Gly Lys Leu Arg Ala Asn His Met Met
380 385 390
Ser Pro Thr Leu Ile Gln Ile Asp Arg Ala Asn Pro Trp Ser Ala
395 400 405
Cys Ser Ala Ala Ile Ile Thr Asp Phe Leu Asp Ser Gly His Gly
410 415 420
Asp Cys Leu Leu Asp Gln Pro Ser Lys Pro Ile Ser Leu Pro Glu
425 430 435
Asp Leu Pro Gly Ala Ser Tyr Thr Leu Ser Gln Gln Cys Glu Leu
440 445 450
Ala Phe Gly Val Gly Ser Lys Pro Cys Pro Tyr Met Gln Tyr Cys
455 460 465
Thr Lys Leu Trp Cys Thr Gly Lys Ala Lys Gly Gln Met Val Cys
470 475 480
Gln Thr Arg His Phe Pro Trp Ala Asp Gly Thr Ser Cys Gly Glu
485 490 495
Gly Lys Leu Cys Leu Lys Gly Ala Cys Val Glu Arg His Asn Leu
500 505 510
Asn Lys His Arg Val Asp Gly Ser Trp Ala Lys Trp Asp Pro Tyr

```
                515                 520                 525
Gly Pro Cys Ser Arg Thr Cys Gly Gly Gly Val Gln Leu Ala Arg
                530                 535                 540
Arg Gln Cys Thr Asn Pro Thr Pro Ala Asn Gly Gly Lys Tyr Cys
                545                 550                 555
Glu Gly Val Arg Val Lys Tyr Arg Ser Cys Asn Leu Glu Pro Cys
                560                 565                 570
Pro Ser Ser Ala Ser Gly Lys Ser Phe Arg Glu Glu Gln Cys Glu
                575                 580                 585
Ala Phe Asn Gly Tyr Asn His Ser Thr Asn Arg Leu Thr Leu Ala
                590                 595                 600
Val Ala Trp Val Pro Lys Tyr Ser Gly Val Ser Pro Arg Asp Lys
                605                 610                 615
Cys Lys Leu Ile Cys Arg Ala Asn Gly Thr Gly Tyr Phe Tyr Val
                620                 625                 630
Leu Ala Pro Lys Val Val Val Asp Gly Thr Leu Cys Ser Pro Asp
                635                 640                 645
Ser Thr Ser Val Cys Val Gln Gly Lys Cys Ile Lys Ala Gly Cys
                650                 655                 660
Asp Gly Asn Leu Gly Ser Lys Lys Arg Phe Asp Lys Cys Gly Val
                665                 670                 675
Cys Gly Gly Asp Asn Lys Ser Cys Lys Lys Val Thr Gly Leu Phe
                680                 685                 690
Thr Lys Pro Met His Gly Tyr Asn Phe Val Val Ala Ile Pro Ala
                695                 700                 705
Gly Ala Ser Ser Ile Asp Ile Arg Gln Arg Gly Tyr Lys Gly Leu
                710                 715                 720
Ile Gly Asp Asp Asn Tyr Leu Ala Leu Lys Asn Ser Gln Gly Lys
                725                 730                 735
Tyr Leu Leu Asn Gly His Phe Val Val Ser Ala Val Glu Arg Asp
                740                 745                 750
Leu Val Val Lys Gly Ser Leu Leu Arg Tyr Ser Gly Thr Gly Thr
                755                 760                 765
Ala Val Glu Ser Leu Gln Ala Ser Arg Pro Ile Leu Glu Pro Leu
                770                 775                 780
Thr Val Glu Val Leu Ser Val Gly Lys Met Thr Pro Pro Arg Val
                785                 790                 795
Arg Tyr Ser Phe Tyr Leu Pro Lys Glu Pro Arg Glu Asp Lys Ser
                800                 805                 810
Ser His Pro Pro His Pro Arg Gly Gly Pro Ser Val Leu His Asn
                815                 820                 825
Ser Val Leu Ser Leu Ser Asn Gln Val Glu Gln Pro Asp Asp Arg
                830                 835                 840
Pro Pro Ala Arg Trp Val Ala Gly Ser Trp Gly Pro Cys Ser Ala
                845                 850                 855
Ser Cys Gly Ser Gly Leu Gln Lys Arg Ala Val Asp Trp Arg Gly
                860                 865                 870
Ser Ala Gly Gln Arg Thr Val Pro Ala Cys Asp Ala Ala His Arg
                875                 880                 885
Pro Val Glu Thr Gln Ala Cys Gly Glu Pro Cys Pro Thr Trp Glu
                890                 895                 900
Leu Ser Ala Trp Ser Pro Cys Ser Lys Ser Cys Gly Arg Gly Phe
                905                 910                 915
Gln Arg Arg Ser Leu Lys Cys Val Gly His Gly Gly Arg Leu Leu
                920                 925                 930
Ala Arg Asp Gln Cys Asn Leu His Arg Lys Pro Gln Glu Leu Asp
                935                 940                 945
Phe Cys Val Leu Arg Pro Cys
                950
```

<210> 12
<211> 898
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 7604035CD1

<400> 12
Met Glu Asn Trp Thr Gly Arg Pro Trp Leu Tyr Leu Leu Leu Leu
1 5 10 15
Leu Ser Leu Pro Gln Leu Cys Leu Asp Gln Glu Val Leu Ser Gly
20 25 30
His Ser Leu Gln Thr Pro Thr Glu Glu Gly Gln Gly Pro Glu Gly
35 40 45
Val Trp Gly Pro Trp Val Gln Trp Ala Ser Cys Ser Gln Pro Cys
50 55 60
Gly Val Gly Val Gln Arg Arg Ser Arg Thr Cys Gln Leu Pro Thr
65 70 75
Val Gln Leu His Pro Ser Leu Pro Leu Pro Pro Arg Pro Pro Arg
80 85 90
His Pro Glu Ala Leu Leu Pro Arg Gly Gln Gly Pro Arg Pro Gln
95 100 105
Thr Ser Pro Glu Thr Leu Pro Leu Tyr Arg Thr Gln Ser Arg Gly
110 115 120
Arg Gly Gly Pro Leu Arg Gly Pro Ala Ser His Leu Gly Arg Glu
125 130 135
Glu Thr Gln Glu Ile Arg Ala Ala Arg Arg Ser Arg Leu Arg Asp
140 145 150
Pro Ile Lys Pro Gly Met Phe Gly Tyr Gly Arg Val Pro Phe Ala
155 160 165
Leu Pro Leu His Arg Asn Arg Arg His Pro Arg Ser Pro Pro Arg
170 175 180
Ser Glu Leu Ser Leu Ile Ser Ser Arg Gly Glu Glu Pro Ile Pro
185 190 195
Ser Pro Thr Pro Arg Ala Glu Pro Phe Ser Ala Asn Gly Ser Pro
200 205 210
Gln Thr Glu Leu Pro Pro Thr Glu Leu Ser Val His Thr Pro Ser
215 220 225
Pro Gln Ala Glu Pro Leu Ser Pro Glu Thr Ala Gln Thr Glu Val
230 235 240
Ala Pro Arg Thr Arg Pro Ala Pro Leu Arg His His Pro Arg Ala
245 250 255
Gln Ala Ser Gly Thr Glu Pro Pro Ser Pro Thr His Ser Leu Gly
260 265 270
Glu Gly Gly Phe Phe Arg Ala Ser Pro Gln Pro Arg Arg Pro Ser
275 280 285
Ser Gln Gly Trp Ala Ser Pro Gln Val Ala Gly Arg Arg Pro Asp
290 295 300
Pro Phe Pro Ser Val Pro Arg Gly Arg Gly Gln Gln Gly Gln Gly
305 310 315
Pro Trp Gly Thr Gly Gly Thr Pro His Gly Pro Arg Leu Glu Pro
320 325 330
Asp Pro Gln His Pro Gly Ala Trp Leu Pro Leu Leu Ser Asn Gly
335 340 345
Pro His Ala Ser Ser Leu Trp Ser Leu Phe Ala Pro Ser Ser Pro
350 355 360
Ile Pro Arg Cys Ser Gly Glu Ser Glu Gln Leu Arg Ala Cys Ser
365 370 375
Gln Ala Pro Cys Pro Pro Glu Gln Pro Asp Pro Arg Ala Leu Gln
380 385 390
Cys Ala Ala Phe Asn Ser Gln Glu Phe Met Gly Gln Leu Tyr Gln
395 400 405
Trp Glu Pro Phe Thr Glu Val Gln Gly Ser Gln Arg Cys Glu Leu
410 415 420
Asn Cys Arg Pro Arg Gly Phe Arg Phe Tyr Val Arg His Thr Glu
425 430 435
Lys Val Gln Asp Gly Thr Leu Cys Gln Pro Gly Ala Pro Asp Ile
440 445 450
Cys Val Ala Gly Arg Cys Leu Ser Pro Gly Cys Asp Gly Ile Leu
455 460 465
Gly Ser Gly Arg Arg Pro Asp Gly Cys Gly Val Cys Gly Gly Asp
470 475 480

Asp Ser Thr Cys Arg Leu Val Ser Gly Asn Leu Thr Asp Arg Gly
485 490 495
Gly Pro Leu Gly Tyr Gln Lys Ile Leu Trp Ile Pro Ala Gly Ala
500 505 510
Leu Arg Leu Gln Ile Ala Gln Leu Arg Pro Ser Ser Asn Tyr Leu
515 520 525
Ala Leu Arg Gly Pro Gly Gly Arg Ser Ile Ile Asn Gly Asn Trp
530 535 540
Ala Val Asp Pro Pro Gly Ser Tyr Arg Ala Gly Gly Thr Val Phe
545 550 555
Arg Tyr Asn Arg Pro Pro Arg Glu Glu Gly Lys Gly Glu Ser Leu
560 565 570
Ser Ala Glu Gly Pro Thr Thr Gln Pro Val Asp Val Tyr Met Ile
575 580 585
Phe Gln Glu Glu Asn Pro Gly Val Phe Tyr Gln Tyr Val Ile Ser
590 595 600
Ser Pro Pro Pro Ile Leu Glu Asn Pro Thr Pro Glu Pro Pro Val
605 610 615
Pro Gln Leu Gln Pro Glu Ile Leu Arg Val Glu Pro Pro Leu Ala
620 625 630
Pro Ala Pro Arg Pro Ala Arg Thr Pro Gly Thr Leu Gln Arg Gln
635 640 645
Val Arg Ile Pro Gln Met Pro Ala Pro Pro His Pro Arg Thr Pro
650 655 660
Leu Gly Ser Pro Ala Ala Tyr Trp Lys Arg Val Gly His Ser Ala
665 670 675
Cys Ser Ala Ser Cys Gly Lys Gly Val Trp Arg Pro Ile Phe Leu
680 685 690
Cys Ile Ser Arg Glu Ser Gly Glu Glu Leu Asp Glu Arg Ser Cys
695 700 705
Ala Ala Gly Ala Arg Pro Pro Ala Ser Pro Glu Pro Cys His Gly
710 715 720
Thr Pro Cys Pro Pro Tyr Trp Glu Ala Gly Glu Trp Thr Ser Cys
725 730 735
Ser Arg Ser Cys Gly Pro Gly Thr Gln His Arg Gln Leu Gln Cys
740 745 750
Arg Gln Glu Phe Gly Gly Gly Gly Ser Ser Val Pro Pro Glu Arg
755 760 765
Cys Gly His Leu Pro Arg Pro Asn Ile Thr Gln Ser Cys Gln Leu
770 775 780
Arg Leu Cys Gly His Trp Glu Val Gly Ser Pro Trp Ser Gln Cys
785 790 795
Ser Val Arg Cys Gly Arg Gly Gln Arg Ser Arg Gln Val Arg Cys
800 805 810
Val Gly Asn Asn Gly Asp Glu Val Ser Glu Gln Glu Cys Ala Ser
815 820 825
Gly Pro Pro Gln Pro Pro Ser Arg Glu Ala Cys Asp Met Gly Pro
830 835 840
Cys Thr Thr Ala Trp Phe His Ser Asp Trp Ser Ser Lys Cys Ser
845 850 855
Ala Glu Cys Gly Thr Gly Ile Gln Arg Arg Ser Val Val Cys Leu
860 865 870
Gly Ser Gly Ala Ala Thr Arg Ala Arg Pro Gly Gly Ser Arg Ser
875 880 885
Arg Asn Trp Ala Glu Leu Ser Asn Arg Lys Pro Ala Pro
890 895

<210> 13
<211> 631
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3473847CD1

<400> 13

Met Phe Leu Leu Ala Trp Gly Gln Asp Pro Trp Arg Leu Pro Gly
1 5 10 15

Thr Tyr Val Val Val Leu Lys Glu Glu Thr His Leu Ser Gln Ser
20 25 30

Glu Arg Thr Ala Arg Arg Leu Gln Ala Gln Ala Ala Arg Arg Gly
35 40 45

Tyr Leu Thr Lys Ile Leu His Val Phe His Gly Leu Leu Pro Gly
50 55 60

Phe Leu Val Lys Met Ser Gly Asp Leu Leu Glu Leu Ala Leu Lys
65 70 75

Leu Pro His Val Asp Tyr Ile Glu Glu Asp Ser Ser Val Phe Ala
80 85 90

Gln Ser Ile Pro Trp Asn Leu Glu Arg Ile Thr Pro Pro Arg Tyr
95 100 105

Arg Ala Asp Glu Tyr Gln Pro Pro Asp Gly Gly Ser Leu Val Glu
110 115 120

Val Tyr Leu Leu Asp Thr Ser Ile Gln Ser Asp His Arg Glu Ile
125 130 135

Glu Gly Arg Val Met Val Thr Asp Phe Glu Asn Val Pro Glu Glu
140 145 150

Asp Gly Thr Arg Phe His Arg Gln Ala Ser Lys Cys Asp Ser His
155 160 165

Gly Thr His Leu Ala Gly Val Val Ser Gly Arg Asp Ala Gly Val
170 175 180

Ala Lys Gly Ala Ser Met Arg Ser Leu Arg Val Leu Asn Cys Gln
185 190 195

Gly Lys Gly Thr Val Ser Gly Thr Leu Ile Gly Leu Glu Phe Ile
200 205 210

Arg Lys Ser Gln Leu Val Gln Pro Val Gly Pro Leu Val Val Leu
215 220 225

Leu Pro Leu Ala Gly Gly Tyr Ser Arg Val Leu Asn Ala Ala Cys
230 235 240

Gln Arg Leu Ala Arg Ala Gly Val Val Leu Val Thr Ala Ala Gly
245 250 255

Asn Phe Arg Asp Asp Ala Cys Leu Tyr Ser Pro Ala Ser Ala Pro
260 265 270

Glu Val Ile Thr Val Gly Ala Thr Asn Ala Gln Asp Gln Pro Val
275 280 285

Thr Leu Gly Thr Leu Gly Thr Asn Phe Gly Arg Cys Val Asp Leu
290 295 300

Phe Ala Pro Gly Glu Asp Ile Ile Gly Ala Ser Ser Asp Cys Ser
305 310 315

Thr Cys Phe Val Ser Gln Ser Gly Thr Ser Gln Ala Ala Ala His
320 325 330

Val Ala Gly Ile Ala Ala Met Met Leu Ser Ala Glu Pro Glu Leu
335 340 345

Thr Leu Ala Glu Leu Arg Gln Arg Leu Ile His Phe Ser Ala Lys
350 355 360

Asp Val Ile Asn Glu Ala Trp Phe Pro Glu Asp Gln Arg Val Leu
365 370 375

Thr Pro Asn Leu Val Ala Ala Leu Pro Pro Ser Thr His Gly Ala
380 385 390

Gly Trp Gln Leu Phe Cys Arg Thr Val Trp Ser Ala His Ser Gly
395 400 405

Pro Thr Arg Met Ala Thr Ala Ile Ala Arg Cys Ala Pro Asp Glu
410 415 420

Glu Leu Leu Ser Cys Ser Ser Phe Ser Arg Ser Gly Lys Arg Arg
425 430 435

Gly Glu Arg Met Glu Ala Gln Gly Gly Lys Leu Val Cys Arg Ala
440 445 450

His Asn Ala Phe Gly Gly Glu Gly Val Tyr Ala Ile Ala Arg Cys
455 460 465

Cys Leu Leu Pro Gln Ala Asn Cys Ser Val His Thr Ala Pro Pro
470 475 480

Ala Glu Ala Ser Met Gly Thr Arg Val His Cys His Gln Gln Gly
485 490 495

His Val Leu Thr Gly Cys Ser Ser His Trp Glu Val Glu Asp Leu

```
                    500                     505                     510
Gly Thr His Lys Pro Pro Val Leu Arg Pro Arg Gly Gln Pro Asn
                    515                     520                     525
Gln Cys Val Gly His Arg Glu Ala Ser Ile His Ala Ser Cys Cys
                    530                     535                     540
His Ala Pro Gly Leu Glu Cys Lys Val Lys Glu His Gly Ile Pro
                    545                     550                     555
Ala Pro Gln Glu Gln Val Thr Val Ala Cys Glu Glu Gly Trp Thr
                    560                     565                     570
Leu Thr Gly Cys Ser Ala Leu Pro Gly Thr Ser His Val Leu Gly
                    575                     580                     585
Ala Tyr Ala Val Asp Asn Thr Cys Val Val Arg Ser Arg Asp Val
                    590                     595                     600
Ser Thr Thr Gly Ser Thr Ser Glu Glu Ala Val Thr Ala Val Ala
                    605                     610                     615
Ile Cys Cys Arg Ser Arg His Leu Ala Gln Ala Ser Gln Glu Leu
                    620                     625                     630
Gln


<210> 14
<211> 470
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3750004CD1

<400> 14
Met Arg His Arg Thr Asp Leu Gly Gln Asn Leu Leu Leu Phe Leu
 1               5                  10                  15
Trp Ala Leu Leu Asn Cys Gly Leu Gly Val Ser Ala Gln Gly Pro
                20                  25                  30
Gly Glu Trp Thr Pro Trp Val Ser Trp Thr Arg Cys Ser Ser Ser
                35                  40                  45
Cys Gly Arg Gly Val Ser Val Arg Ser Arg Arg Cys Leu Arg Leu
                50                  55                  60
Pro Gly Glu Glu Pro Cys Trp Gly Asp Ser His Glu Tyr Arg Leu
                65                  70                  75
Cys Gln Leu Pro Asp Cys Pro Pro Gly Ala Val Pro Phe Arg Asp
                80                  85                  90
Leu Gln Cys Ala Leu Tyr Asn Gly Arg Pro Val Leu Gly Thr Gln
                95                 100                 105
Lys Thr Tyr Gln Trp Val Pro Phe His Gly Ala Pro Asn Gln Cys
               110                 115                 120
Asp Leu Asn Cys Leu Ala Glu Gly His Ala Phe Tyr His Ser Phe
               125                 130                 135
Gly Arg Val Leu Asp Gly Thr Ala Cys Ser Pro Gly Ala Gln Gly
               140                 145                 150
Val Cys Val Ala Gly Arg Cys Leu Ser Ala Gly Cys Asp Gly Leu
               155                 160                 165
Leu Gly Ser Gly Ala Leu Glu Asp Arg Cys Gly Arg Cys Gly Gly
               170                 175                 180
Ala Asn Asp Ser Cys Leu Phe Val Gln Arg Val Phe Arg Asp Ala
               185                 190                 195
Gly Ala Phe Ala Gly Tyr Trp Asn Val Thr Leu Ile Pro Glu Gly
               200                 205                 210
Ala Arg His Ile Arg Val Glu His Arg Ser Arg Asn His Leu Gly
               215                 220                 225
Ile Leu Gly Ser Leu Met Gly Gly Asp Gly Arg Tyr Val Leu Asn
               230                 235                 240
Gly His Trp Val Val Ser Pro Pro Gly Thr Tyr Glu Ala Ala Gly
               245                 250                 255
Thr His Val Val Tyr Thr Arg Asp Thr Gly Pro Gln Glu Thr Leu
               260                 265                 270
Gln Ala Ala Gly Pro Thr Ser His Asp Leu Leu Leu Gln Val Leu
```

```
                275                 280                 285
Leu Gln Glu Pro Asn Pro Gly Ile Glu Phe Glu Phe Trp Leu Pro
                290                 295                 300
Arg Glu Arg Tyr Ser Pro Phe Gln Ala Arg Val Gln Ala Leu Gly
                305                 310                 315
Trp Pro Leu Arg Gln Pro Gln Pro Arg Gly Val Glu Pro Gln Pro
                320                 325                 330
Pro Ala Ala Pro Ala Val Thr Pro Ala Gln Thr Pro Thr Leu Ala
                335                 340                 345
Pro Asp Pro Cys Pro Pro Cys Pro Asp Thr Arg Gly Arg Ala His
                350                 355                 360
Arg Leu Leu His Tyr Cys Gly Ser Asp Phe Val Phe Gln Ala Arg
                365                 370                 375
Val Leu Gly His His His Gln Ala Gln Glu Thr Arg Tyr Glu Val
                380                 385                 390
Arg Ile Gln Leu Val Tyr Lys Asn Arg Ser Pro Leu Arg Ala Arg
                395                 400                 405
Glu Tyr Val Trp Ala Pro Gly His Cys Pro Cys Pro Met Leu Ala
                410                 415                 420
Pro His Arg Asp Tyr Leu Met Ala Val Gln Arg Leu Val Ser Pro
                425                 430                 435
Asp Gly Thr Gln Asp Gln Leu Leu Leu Pro His Ala Gly Tyr Ala
                440                 445                 450
Arg Pro Trp Ser Pro Ala Glu Asp Ser Arg Ile Arg Leu Thr Ala
                455                 460                 465
Arg Arg Cys Pro Gly
                470
```

<210> 15
<211> 110
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4904126CD1

<400> 15

```
Met Ala Asp Lys Val Leu Lys Glu Lys Arg Lys Gln Phe Ile Arg
 1               5                  10                  15
Ser Val Gly Glu Gly Thr Ile Asn Gly Leu Leu Gly Glu Leu Leu
                20                  25                  30
Glu Thr Arg Val Leu Ser Gln Glu Glu Ile Glu Ile Val Lys Cys
                35                  40                  45
Glu Asn Ala Thr Val Met Asp Lys Ala Arg Ala Leu Leu Asp Ser
                50                  55                  60
Val Ile Arg Lys Gly Ala Pro Ala Cys Gln Ile Cys Ile Thr Tyr
                65                  70                  75
Ile Cys Glu Glu Asp Ser His Leu Ala Gly Thr Leu Gly Leu Ser
                80                  85                  90
Ala Gly Pro Thr Ser Gly Asn His Leu Thr Thr Gln Asp Ser Gln
                95                 100                 105
Ile Val Leu Pro Ser
               110
```

<210> 16
<211> 879
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71268415CD1

<400> 16

```
Met Ser Leu Phe Ile Phe Cys Arg Gln Leu Phe Ala Pro Ser Tyr
 1               5                  10                  15
```

Thr Glu Thr His Tyr Thr Ser Ser Gly Asn Pro Gln Thr Thr Thr
20 25 30
Arg Lys Leu Glu Asp His Cys Phe Tyr His Gly Thr Val Arg Glu
35 40 45
Thr Glu Leu Ser Ser Val Thr Leu Ser Thr Cys Arg Gly Ile Arg
50 55 60
Gly Leu Ile Thr Val Ser Ser Asn Leu Ser Tyr Val Ile Glu Pro
65 70 75
Leu Pro Asp Ser Lys Gly Gln His Leu Ile Tyr Arg Ser Glu His
80 85 90
Leu Lys Pro Pro Pro Gly Asn Cys Gly Phe Glu His Ser Lys Pro
95 100 105
Thr Thr Arg Asp Trp Ala Leu Gln Phe Thr Gln Gln Thr Lys Lys
110 115 120
Arg Pro Arg Arg Met Lys Arg Glu Asp Leu Asn Ser Met Lys Tyr
125 130 135
Val Glu Leu Tyr Leu Val Ala Asp Tyr Leu Glu Phe Gln Lys Asn
140 145 150
Arg Arg Asp Gln Asp Ala Thr Lys His Lys Leu Ile Glu Ile Ala
155 160 165
Asn Tyr Val Asp Lys Phe Tyr Arg Ser Leu Asn Ile Arg Ile Ala
170 175 180
Leu Val Gly Leu Glu Val Trp Thr His Gly Asn Met Cys Glu Val
185 190 195
Ser Glu Asn Pro Tyr Ser Thr Leu Trp Ser Phe Leu Ser Trp Arg
200 205 210
Arg Lys Leu Leu Ala Gln Lys Tyr His Asp Asn Ala Gln Leu Ile
215 220 225
Thr Gly Met Ser Phe His Gly Thr Thr Ile Gly Leu Ala Pro Leu
230 235 240
Met Ala Met Cys Ser Val Tyr Gln Ser Gly Gly Val Asn Met Asp
245 250 255
His Ser Glu Asn Ala Ile Gly Val Ala Ala Thr Met Ala His Glu
260 265 270
Met Gly His Asn Phe Gly Met Thr His Asp Ser Ala Asp Cys Cys
275 280 285
Ser Ala Ser Ala Ala Asp Gly Gly Cys Ile Met Ala Ala Ala Thr
290 295 300
Gly His Pro Phe Pro Lys Val Phe Asn Gly Cys Asn Arg Arg Glu
305 310 315
Leu Asp Arg Tyr Leu Gln Ser Gly Gly Gly Met Cys Leu Ser Asn
320 325 330
Met Pro Asp Thr Arg Met Leu Tyr Gly Gly Arg Arg Cys Gly Asn
335 340 345
Gly Tyr Leu Glu Asp Gly Glu Glu Cys Asp Cys Gly Glu Glu Glu
350 355 360
Glu Cys Asn Asn Pro Cys Cys Asn Ala Ser Asn Cys Thr Leu Arg
365 370 375
Pro Gly Ala Glu Cys Ala His Gly Ser Cys Cys His Gln Cys Lys
380 385 390
Leu Leu Ala Pro Gly Thr Leu Cys Arg Glu Gln Ala Arg Gln Cys
395 400 405
Asp Leu Pro Glu Phe Cys Thr Gly Lys Ser Pro His Cys Pro Thr
410 415 420
Asn Phe Tyr Gln Met Asp Gly Thr Pro Cys Glu Gly Gly Gln Ala
425 430 435
Tyr Cys Tyr Asn Gly Met Cys Leu Thr Tyr Gln Glu Gln Cys Gln
440 445 450
Gln Leu Trp Gly Pro Gly Ala Arg Pro Ala Pro Asp Leu Cys Phe
455 460 465
Glu Lys Val Asn Val Ala Gly Asp Thr Phe Gly Asn Cys Gly Lys
470 475 480
Asp Met Asn Gly Glu His Arg Lys Cys Asn Met Arg Asp Ala Lys
485 490 495
Cys Gly Lys Ile Gln Cys Gln Ser Ser Glu Ala Arg Pro Leu Glu
500 505 510
Ser Asn Ala Val Pro Ile Asp Thr Thr Ile Ile Met Asn Gly Arg

```
                515                     520                     525
Gln Ile Gln Cys Arg Gly Thr His Val Tyr Arg Gly Pro Glu Glu
                530                     535                     540
Glu Gly Asp Met Leu Asp Pro Gly Leu Val Met Thr Gly Thr Lys
                545                     550                     555
Cys Gly Tyr Asn His Ile Cys Phe Glu Gly Gln Cys Arg Asn Thr
                560                     565                     570
Ser Phe Phe Glu Thr Glu Gly Cys Gly Lys Lys Cys Asn Gly His
                575                     580                     585
Gly Val Cys Asn Asn Asn Gln Asn Cys His Cys Leu Pro Gly Trp
                590                     595                     600
Ala Pro Pro Phe Cys Asn Thr Pro Gly His Gly Gly Ser Ile Asp
                605                     610                     615
Ser Gly Pro Met Pro Pro Glu Ser Val Gly Pro Val Val Ala Gly
                620                     625                     630
Val Leu Val Ala Ile Leu Val Leu Ala Val Leu Met Leu Met Tyr
                635                     640                     645
Tyr Cys Cys Arg Gln Asn Asn Lys Leu Gly Gln Leu Lys Pro Ser
                650                     655                     660
Ala Leu Pro Ser Lys Leu Arg Gln Gln Phe Ser Cys Pro Phe Arg
                665                     670                     675
Val Ser Gln Asn Ser Gly Thr Gly His Ala Asn Pro Thr Phe Lys
                680                     685                     690
Leu Gln Thr Pro Gln Gly Lys Arg Lys Val Ile Asn Thr Pro Glu
                695                     700                     705
Ile Leu Arg Lys Pro Ser Gln Pro Pro Pro Arg Pro Pro Pro Asp
                710                     715                     720
Tyr Leu Arg Gly Gly Ser Pro Pro Ala Pro Leu Pro Ala His Leu
                725                     730                     735
Ser Arg Ala Ala Arg Asn Ser Pro Gly Pro Gly Ser Gln Ile Glu
                740                     745                     750
Arg Thr Glu Ser Ser Arg Arg Pro Pro Pro Ser Arg Pro Ile Pro
                755                     760                     765
Pro Ala Pro Asn Cys Ile Val Ser Gln Asp Phe Ser Arg Pro Arg
                770                     775                     780
Pro Pro Gln Lys Ala Leu Pro Ala Asn Pro Val Pro Gly Arg Arg
                785                     790                     795
Ser Leu Pro Arg Pro Gly Gly Ala Ser Pro Leu Arg Pro Pro Gly
                800                     805                     810
Ala Gly Pro Gln Gln Ser Arg Pro Leu Ala Ala Leu Ala Pro Lys
                815                     820                     825
Val Ser Pro Arg Glu Ala Leu Lys Val Lys Ala Gly Thr Arg Gly
                830                     835                     840
Leu Gln Gly Gly Arg Cys Arg Val Glu Lys Thr Lys Gln Phe Met
                845                     850                     855
Leu Leu Val Val Trp Thr Glu Leu Pro Glu Gln Lys Pro Arg Ala
                860                     865                     870
Lys His Ser Cys Phe Leu Val Pro Ala
                875

&lt;210&gt; 17
&lt;211&gt; 850
&lt;212&gt; PRT
&lt;213&gt; Homo sapiens

&lt;220&gt;
&lt;221&gt; misc_feature
&lt;223&gt; Incyte ID No: 7473301CD1

&lt;400&gt; 17
Met Asp Lys Glu Asn Ser Asp Val Ser Ala Ala Pro Ala Asp Leu
 1               5                  10                  15
Lys Ile Ser Asn Ile Ser Val Gln Val Val Ser Ala Gln Lys Lys
                20                  25                  30
Leu Pro Val Arg Arg Pro Pro Leu Pro Gly Arg Arg Leu Pro Leu
                35                  40                  45
Pro Gly Arg Arg Pro Pro Gln Arg Pro Ile Gly Lys Ala Lys Pro
```

```
                    50                  55                  60
Lys Lys Gln Ser Lys Lys Lys Val Pro Phe Trp Asn Val Gln Asn
                65                  70                  75
Lys Ile Ile Leu Phe Thr Val Phe Leu Phe Ile Leu Ala Val Ile
                80                  85                  90
Ala Trp Thr Leu Leu Trp Leu Tyr Ile Ser Lys Thr Glu Ser Lys
                95                  100                 105
Asp Ala Phe Tyr Phe Ala Gly Met Phe Arg Ile Thr Asn Ile Glu
                110                 115                 120
Phe Leu Pro Glu Tyr Arg Gln Lys Glu Ser Arg Glu Phe Leu Ser
                125                 130                 135
Val Ser Arg Thr Val Gln Gln Val Ile Asn Leu Val Tyr Thr Thr
                140                 145                 150
Ser Ala Phe Ser Lys Phe Tyr Glu Gln Ser Val Val Ala Asp Val
                155                 160                 165
Ser Ser Asn Asn Lys Gly Gly Leu Leu Val His Phe Trp Ile Val
                170                 175                 180
Phe Val Met Pro Arg Ala Lys Gly His Ile Phe Cys Glu Asp Cys
                185                 190                 195
Val Ala Ala Ile Leu Lys Asp Ser Ile Gln Thr Ser Ile Ile Asn
                200                 205                 210
Arg Thr Ser Val Gly Ser Leu Gln Gly Leu Ala Val Asp Met Asp
                215                 220                 225
Ser Val Val Leu Asn Gly Asp Cys Trp Ser Phe Leu Lys Lys Lys
                230                 235                 240
Lys Arg Lys Glu Asn Gly Ala Val Ser Thr Asp Lys Gly Cys Ser
                245                 250                 255
Gln Tyr Phe Tyr Ala Glu His Leu Ser Leu His Tyr Pro Leu Glu
                260                 265                 270
Ile Ser Ala Ala Ser Gly Arg Leu Met Cys His Phe Lys Leu Val
                275                 280                 285
Ala Ile Val Gly Tyr Leu Ile Arg Leu Ser Ile Lys Ser Ile Gln
                290                 295                 300
Ile Glu Ala Asp Asn Cys Val Thr Asp Ser Leu Thr Ile Tyr Asp
                305                 310                 315
Ser Leu Leu Pro Ile Arg Ser Ser Ile Leu Tyr Arg Ile Cys Glu
                320                 325                 330
Pro Thr Arg Thr Leu Met Ser Phe Val Ser Thr Asn Asn Leu Met
                335                 340                 345
Leu Val Thr Phe Lys Ser Pro His Ile Arg Arg Leu Ser Gly Ile
                350                 355                 360
Arg Ala Tyr Phe Glu Val Ile Pro Glu Gln Lys Cys Glu Asn Thr
                365                 370                 375
Val Leu Val Lys Asp Ile Thr Gly Phe Glu Gly Lys Ile Ser Ser
                380                 385                 390
Pro Tyr Tyr Pro Ser Tyr Tyr Pro Pro Lys Cys Lys Cys Thr Trp
                395                 400                 405
Lys Phe Gln Thr Ser Leu Ser Thr Leu Gly Ile Ala Leu Lys Phe
                410                 415                 420
Tyr Asn Tyr Ser Ile Thr Lys Lys Ser Met Lys Gly Cys Glu His
                425                 430                 435
Gly Trp Trp Glu Ile Tyr Glu His Met Tyr Cys Gly Ser Tyr Met
                440                 445                 450
Asp His Gln Thr Ile Phe Arg Val Pro Ser Pro Leu Val His Ile
                455                 460                 465
Gln Leu Gln Cys Ser Ser Arg Leu Ser Gly Lys Pro Leu Leu Ala
                470                 475                 480
Glu Tyr Gly Ser Tyr Asn Ile Ser Gln Pro Cys Pro Val Gly Ser
                485                 490                 495
Phe Arg Cys Ser Ser Gly Leu Cys Val Pro Gln Ala Gln Arg Gly
                500                 505                 510
Asp Gly Val Asn Asp Cys Phe Asp Glu Ser Asp Glu Leu Phe Cys
                515                 520                 525
Val Ser Pro Gln Pro Ala Cys Asn Thr Ser Ser Phe Arg Gln His
                530                 535                 540
Gly Pro Leu Ile Cys Asp Gly Phe Arg Asp Cys Glu Asn Gly Arg
                545                 550                 555
```

Asp Glu Gln Asn Cys Thr Gln Ser Ile Pro Cys Asn Asn Arg Thr
560 565 570
Phe Lys Cys Gly Asn Asp Ile Cys Phe Arg Lys Gln Asn Ala Lys
575 580 585
Cys Asp Gly Thr Val Asp Cys Pro Asp Gly Ser Asp Glu Glu Gly
590 595 600
Cys Thr Cys Ser Arg Ser Ser Ser Ala Leu His Arg Ile Ile Gly
605 610 615
Gly Thr Asp Thr Leu Glu Gly Gly Trp Pro Trp Gln Val Ser Leu
620 625 630
His Phe Val Gly Ser Ala Tyr Cys Gly Ala Ser Val Ile Ser Arg
635 640 645
Glu Trp Leu Leu Ser Ala Ala His Cys Phe His Gly Asn Arg Leu
650 655 660
Ser Asp Pro Thr Pro Trp Thr Ala His Leu Gly Met Tyr Val Gln
665 670 675
Gly Asn Ala Lys Phe Val Ser Pro Val Arg Arg Ile Val Val His
680 685 690
Glu Tyr Tyr Asn Ser Gln Thr Phe Asp Tyr Asp Ile Ala Leu Leu
695 700 705
Gln Leu Ser Ile Ala Trp Pro Glu Thr Leu Lys Gln Leu Ile Gln
710 715 720
Pro Ile Cys Ile Pro Pro Thr Gly Gln Arg Val Arg Ser Gly Glu
725 730 735
Lys Cys Trp Val Thr Gly Trp Gly Arg Arg His Glu Ala Asp Asn
740 745 750
Lys Gly Ser Leu Val Leu Gln Gln Ala Glu Val Glu Leu Ile Asp
755 760 765
Gln Thr Leu Cys Val Ser Thr Tyr Gly Ile Ile Thr Ser Arg Met
770 775 780
Leu Cys Ala Gly Ile Met Ser Gly Lys Arg Asp Ala Cys Lys Gly
785 790 795
Asp Ser Gly Gly Pro Leu Ser Cys Arg Arg Lys Ser Asp Gly Lys
800 805 810
Trp Ile Leu Thr Gly Ile Val Ser Trp Gly His Gly Cys Gly Arg
815 820 825
Pro Asn Phe Pro Gly Val Tyr Thr Arg Val Ser Asn Phe Val Pro
830 835 840
Trp Ile His Lys Tyr Val Pro Ser Leu Leu
845 850

<210> 18
<211> 254
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473308CD1

<400> 18
Met Gln Asp His Arg Lys Gly Lys Ala Ala Val Gly Val Ser Phe
1 5 10 15
Asp Asp Asp Asp Lys Ile Val Gly Gly Tyr Asn Cys Glu Glu Asn
20 25 30
Ser Val Pro Tyr Gln Val Ser Leu Asn Ser Gly Tyr His Phe Cys
35 40 45
Val Gly Ser Leu Asn Arg Glu Tyr Cys Ile Gln Val Arg Leu Gly
50 55 60
Glu His Asn Ile Glu Val Leu Glu Gly Asn Glu Gln Phe Ile Tyr
65 70 75
Ala Val Lys Ile Ile Arg His Pro Lys Tyr Asn Ser Trp Thr Leu
80 85 90
Asp Asn Asp Ile Leu Leu Ile Lys Leu Ser Thr Pro Ala Ile Ile
95 100 105
Asn Ala His Val Ser Thr Ile Ser Leu Pro Thr Thr Pro Pro Ala
110 115 120

```
Ala Gly Thr Glu Cys Leu Ile Ser Gly Trp Gly Asn Thr Leu Ser
                125                 130                 135
Ser Gly Ala Asp Tyr Pro Asp Glu Leu Gln Cys Leu Asp Ala Pro
                140                 145                 150
Val Leu Ser Gln Ala Glu Tyr Glu Ala Ser Tyr Pro Gly Lys Ile
                155                 160                 165
Thr Asn Asn Val Phe Cys Val Gly Phe Leu Glu Gly Gly Lys Asp
                170                 175                 180
Ser Cys Gln Ile Ile Pro Ile Lys Val Gln Gln Leu Val Thr Ser
                185                 190                 195
Ser Gln Glu Thr Asp Ile Arg Ile Pro Met Ala Leu Gln Thr Ala
                200                 205                 210
Ala Ser Thr Ser Tyr Leu Gly Pro Leu Asp Ser Leu His Arg Lys
                215                 220                 225
Val Ser His Pro Thr Glu Lys Arg Cys Gln Gln Lys Gln Gly Met
                230                 235                 240
Lys Ile Thr Asp Asn His Gly Ile Thr Ser Lys Trp Ser Val
                245                 250
```

<210> 19
<211> 568
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7478021CD1

<400> 19

```
Met Leu Ala Ala Ser Ile Phe Arg Pro Thr Leu Leu Leu Cys Trp
 1               5                  10                  15
Leu Ala Ala Pro Trp Pro Thr Gln Pro Glu Ser Leu Phe His Ser
                20                  25                  30
Arg Asp Arg Ser Asp Leu Glu Pro Ser Pro Leu Arg Gln Ala Lys
                35                  40                  45
Pro Ile Ala Asp Leu His Ala Ala Gln Arg Phe Leu Ser Arg Tyr
                50                  55                  60
Gly Trp Ser Gly Val Trp Ala Ala Trp Gly Pro Ser Pro Glu Gly
                65                  70                  75
Pro Pro Glu Thr Pro Lys Gly Ala Ala Leu Ala Glu Ala Val Arg
                80                  85                  90
Arg Phe Gln Arg Ala Asn Ala Leu Pro Ala Ser Gly Glu Leu Asp
                95                 100                 105
Ala Ala Thr Leu Ala Ala Met Asn Arg Pro Arg Cys Gly Val Pro
                110                115                 120
Asp Met Arg Pro Pro Pro Pro Ser Ala Pro Pro Ser Pro Pro Gly
                125                130                 135
Pro Pro Pro Arg Ala Arg Ser Arg Arg Ser Pro Arg Ala Pro Leu
                140                145                 150
Ser Leu Ser Arg Arg Gly Trp Gln Pro Arg Gly Tyr Pro Asp Gly
                155                160                 165
Gly Ala Ala Gln Ala Phe Ser Lys Arg Thr Leu Ser Trp Arg Leu
                170                175                 180
Leu Gly Glu Ala Leu Ser Ser Gln Leu Ser Val Ala Asp Gln Arg
                185                190                 195
Arg Ile Glu Ala Leu Ala Phe Arg Met Trp Ser Glu Val Thr Pro
                200                205                 210
Leu Asp Phe Arg Glu Asp Leu Ala Ala Pro Gly Ala Ala Val Asp
                215                220                 225
Ile Lys Leu Gly Phe Gly Arg Arg His Leu Gly Cys Pro Arg Ala
                230                235                 240
Phe Asp Gly Ser Gly Gln Glu Phe Ala His Ala Trp Arg Leu Gly
                245                250                 255
Asp Ile His Phe Asp Asp Asp Glu His Phe Thr Pro Pro Thr Ser
                260                265                 270
Asp Thr Gly Ile Ser Leu Leu Lys Val Ala Val His Glu Ile Gly
                275                280                 285
```

His Val Leu Gly Leu Pro His Thr Tyr Arg Thr Gly Ser Ile Met
290 295 300
Gln Pro Asn Tyr Ile Pro Gln Glu Pro Ala Phe Glu Leu Asp Trp
305 310 315
Ser Asp Arg Lys Ala Ile Gln Lys Leu Tyr Gly Ser Cys Glu Gly
320 325 330
Ser Phe Asp Thr Ala Phe Asp Trp Ile Arg Lys Glu Arg Asn Gln
335 340 345
Tyr Gly Glu Val Met Val Arg Phe Ser Thr Tyr Phe Phe Arg Asn
350 355 360
Ser Trp Tyr Trp Leu Tyr Glu Asn Arg Asn Asn Arg Thr Arg Tyr
365 370 375
Gly Asp Pro Ile Gln Ile Leu Thr Gly Trp Pro Gly Ile Pro Thr
380 385 390
His Asn Ile Asp Ala Phe Val His Ile Trp Thr Trp Lys Arg Asp
395 400 405
Glu Arg Tyr Phe Phe Gln Gly Asn Gln Tyr Trp Arg Tyr Asp Ser
410 415 420
Asp Lys Asp Gln Ala Leu Thr Glu Asp Glu Gln Gly Lys Ser Tyr
425 430 435
Pro Lys Leu Ile Ser Glu Gly Phe Pro Gly Ile Pro Ser Pro Leu
440 445 450
Asp Thr Ala Phe Tyr Asp Arg Arg Gln Lys Leu Ile Tyr Phe Phe
455 460 465
Lys Glu Ser Leu Val Phe Ala Phe Asp Val Asn Arg Asn Arg Val
470 475 480
Leu Asn Ser Tyr Pro Lys Arg Ile Thr Glu Val Phe Pro Ala Val
485 490 495
Ile Pro Gln Asn His Pro Phe Arg Asn Ile Asp Ser Ala Tyr Tyr
500 505 510
Ser Tyr Ala Tyr Asn Ser Ile Phe Phe Phe Lys Gly Asn Ala Tyr
515 520 525
Trp Lys Val Val Asn Asp Lys Asp Lys Gln Gln Asn Ser Trp Leu
530 535 540
Pro Ala Asn Gly Leu Phe Pro Lys Lys Phe Ile Ser Glu Lys Trp
545 550 555
Phe Asp Val Cys Asp Val His Ile Ser Thr Leu Asn Met
560 565

<210> 20
<211> 306
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4333459CD1

<400> 20
Met Ser Leu Lys Met Leu Ile Ser Arg Asn Lys Leu Ile Leu Leu
1 5 10 15
Leu Gly Ile Val Phe Phe Glu Arg Gly Lys Ser Ala Thr Leu Ser
20 25 30
Leu Pro Lys Ala Pro Ser Cys Gly Gln Ser Leu Val Lys Val Gln
35 40 45
Pro Trp Asn Tyr Phe Asn Ile Phe Ser Arg Ile Leu Gly Gly Ser
50 55 60
Gln Val Glu Lys Gly Ser Tyr Pro Trp Gln Val Ser Leu Lys Gln
65 70 75
Arg Gln Lys His Ile Cys Gly Gly Ser Ile Val Ser Pro Gln Trp
80 85 90
Val Ile Thr Ala Ala His Cys Ile Ala Asn Arg Asn Ile Val Ser
95 100 105
Thr Leu Asn Val Thr Ala Gly Glu Tyr Asp Leu Ser Gln Thr Asp
110 115 120
Pro Gly Glu Gln Thr Leu Thr Ile Glu Thr Val Ile Ile His Pro
125 130 135

```
His Phe Ser Thr Lys Lys Pro Met Asp Tyr Asp Ile Ala Leu Leu
                140                 145                 150
Lys Met Ala Gly Ala Phe Gln Phe Gly His Phe Val Gly Pro Ile
                155                 160                 165
Cys Leu Pro Glu Leu Arg Glu Gln Phe Glu Ala Gly Phe Ile Cys
                170                 175                 180
Thr Thr Ala Gly Trp Gly Arg Leu Thr Glu Gly Gly Val Leu Ser
                185                 190                 195
Gln Val Leu Gln Glu Val Asn Leu Pro Ile Leu Thr Trp Glu Glu
                200                 205                 210
Cys Val Ala Ala Leu Leu Thr Leu Lys Arg Pro Ile Ser Gly Lys
                215                 220                 225
Thr Phe Leu Cys Thr Gly Phe Pro Asp Gly Gly Arg Asp Ala Cys
                230                 235                 240
Gln Gly Asp Ser Gly Gly Ser Leu Met Cys Arg Asn Lys Lys Gly
                245                 250                 255
Ala Trp Thr Leu Ala Gly Val Thr Ser Trp Gly Leu Gly Cys Gly
                260                 265                 270
Arg Gly Trp Arg Asn Asn Val Arg Lys Ser Asp Gln Gly Ser Pro
                275                 280                 285
Gly Ile Phe Thr Asp Ile Ser Lys Val Leu Ser Trp Ile His Glu
                290                 295                 300
His Ile Gln Thr Gly Asn
                305

<210> 21
<211> 953
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6817347CD1

<400> 21
Met Thr Leu Leu Ala Pro Trp Tyr Thr Gly Pro Met Ile Pro Met
  1             5                   10                  15
Asp Val Asn Glu Pro Ser Ser Val Thr Thr Ala Pro Thr Leu Ser
                20                  25                  30
Ser Ser Leu Gln His Ile Ser Ser Phe Leu Ala Thr Gly Lys Lys
                35                  40                  45
Leu Ser Leu His Phe Gly His Pro Arg Glu Cys Glu Val Thr Arg
                50                  55                  60
Ile Asp Asp Lys Asn Arg Arg Gly Leu Glu Asp Ser Glu Pro Gly
                65                  70                  75
Ala Lys Leu Phe Asn Asn Asp Gly Val Cys Cys Cys Leu Gln Lys
                80                  85                  90
Arg Gly Pro Val Asn Ile Thr Ser Val Cys Val Ser Pro Arg Thr
                95                  100                 105
Leu Gln Ile Ser Val Phe Val Leu Ser Glu Lys Tyr Glu Gly Ile
                110                 115                 120
Val Lys Phe Glu Ser Asp Glu Leu Pro Phe Gly Val Ile Gly Ser
                125                 130                 135
Asn Ile Gly Asp Ala His Phe Gln Glu Phe Arg Ala Gly Ile Ser
                140                 145                 150
Trp Lys Pro Val Val Asp Pro Asp Asp Pro Ile Pro Gln Phe Pro
                155                 160                 165
Asp Cys Cys Ser Ser Ser Ser Ser Arg Ile Pro Ser Val Ser Val
                170                 175                 180
Leu Val Ala Val Pro Leu Val Ala Gly His Lys Gly Gln Ala Phe
                185                 190                 195
Ile Glu Arg Met Leu Gly Cys Phe Lys Glu Leu Lys Gln Glu Leu
                200                 205                 210
Thr Gln Glu Gly Pro Gly Gly Gly His Pro Arg Ser Ala Trp Pro
                215                 220                 225
Pro Arg Arg His Ala Gln Trp Pro Pro Glu Pro Cys Glu Gln Gly
                230                 235                 240
```

Glu Glu Pro Pro Pro Val Glu Ala Glu Glu Val Glu Glu Ala Glu
245 250 255
Thr Ala Glu Lys Ala Glu Arg Lys Val Glu Ala Glu Ala Lys Val
260 265 270
Glu Gly Lys Ala Glu Ala Ala Gly Lys Ala Glu Ala Ala Gly Lys
275 280 285
Val Asp Ala Thr Glu Lys Val Glu Thr Ala Gly Lys Val Asp Ala
290 295 300
Ala Gly Lys Val Glu Thr Ala Glu Gly Pro Gly Arg Arg Ala Glu
305 310 315
Leu Lys Leu Glu Pro Glu Pro Glu Pro Val Arg Glu Ala Glu Gln
320 325 330
Glu Pro Lys Gln Glu Leu Glu Asp Glu Asn Pro Ala Arg Ser Gly
335 340 345
Gly Gly Gly Asn Ser Asp Glu Val Pro Pro Pro Thr Leu Pro Ser
350 355 360
Asp Pro Pro Arg Pro Pro Asp Pro Ser Pro Arg Arg Ser Arg Ala
365 370 375
Pro Arg Arg Arg Pro Arg Pro Arg Pro Gln Thr Arg Leu Arg Thr
380 385 390
Pro Pro Gln Pro Arg Pro Arg Pro Pro Pro Arg Pro Arg Pro Arg
395 400 405
Arg Gly Pro Gly Gly Gly Cys Leu Asp Val Asp Phe Ala Val Gly
410 415 420
Pro Pro Gly Cys Ser His Val Asn Ser Phe Lys Val Gly Glu Asn
425 430 435
Trp Arg Gln Glu Leu Arg Val Ile Tyr Gln Cys Phe Val Trp Cys
440 445 450
Gly Thr Pro Glu Thr Arg Lys Ser Lys Ala Lys Ser Cys Ile Cys
455 460 465
His Val Cys Gly Thr His Leu Asn Arg Leu His Ser Cys Leu Ser
470 475 480
Cys Val Phe Phe Gly Cys Phe Thr Glu Lys His Ile His Glu His
485 490 495
Ala Glu Thr Lys Gln His Asn Leu Ala Val Asp Leu Tyr Tyr Gly
500 505 510
Gly Ile Tyr Cys Phe Met Cys Lys Asp Tyr Val Tyr Asp Lys Asp
515 520 525
Ile Glu Gln Ile Ala Lys Glu Glu Gln Gly Glu Ala Leu Lys Leu
530 535 540
Gln Ala Ser Thr Ser Thr Glu Val Ser His Gln Gln Cys Ser Val
545 550 555
Pro Gly Leu Gly Glu Lys Phe Pro Thr Trp Glu Thr Thr Lys Pro
560 565 570
Glu Leu Glu Leu Leu Gly His Asn Pro Arg Arg Arg Arg Ile Thr
575 580 585
Ser Ser Phe Thr Ile Gly Leu Arg Gly Leu Ile Asn Leu Gly Asn
590 595 600
Thr Cys Phe Met Asn Cys Ile Val Gln Ala Leu Thr His Thr Pro
605 610 615
Ile Leu Arg Asp Phe Phe Leu Ser Asp Arg His Arg Cys Glu Met
620 625 630
Pro Ser Pro Glu Leu Cys Leu Val Cys Glu Met Ser Ser Leu Phe
635 640 645
Arg Glu Leu Tyr Ser Gly Asn Pro Ser Pro His Val Pro Tyr Lys
650 655 660
Leu Leu His Leu Val Trp Ile His Ala Arg His Leu Ala Gly Tyr
665 670 675
Arg Gln Gln Asp Ala His Glu Phe Leu Ile Ala Ala Leu Asp Val
680 685 690
Leu His Arg His Cys Lys Gly Asp Asp Val Gly Lys Ala Ala Asn
695 700 705
Asn Pro Asn His Cys Asn Cys Ile Ile Asp Gln Ile Phe Thr Gly
710 715 720
Gly Leu Gln Ser Asp Val Thr Cys Gln Ala Cys His Gly Val Ser
725 730 735
Thr Thr Ile Asp Pro Cys Trp Asp Ile Ser Leu Asp Leu Pro Gly

```
                    740                     745                     750
Ser Cys Thr Ser Phe Trp Pro Met Ser Pro Gly Arg Glu Ser Ser
                755                     760                     765
Val Asn Gly Glu Ser His Ile Pro Gly Ile Thr Thr Leu Thr Asp
                770                     775                     780
Cys Leu Arg Arg Phe Thr Arg Pro Glu His Leu Gly Ser Ser Ala
                785                     790                     795
Lys Ile Lys Cys Gly Ser Cys Gln Ser Tyr Gln Glu Ser Thr Lys
                800                     805                     810
Gln Leu Thr Met Asn Lys Leu Pro Val Val Ala Cys Phe His Phe
                815                     820                     825
Lys Arg Phe Glu His Ser Ala Lys Gln Arg Arg Lys Ile Thr Thr
                830                     835                     840
Tyr Ile Ser Phe Pro Leu Glu Leu Asp Met Thr Pro Phe Met Ala
                845                     850                     855
Ser Ser Lys Glu Ser Arg Met Asn Gly Gln Leu Gln Leu Pro Thr
                860                     865                     870
Asn Ser Gly Asn Asn Glu Asn Lys Tyr Ser Leu Phe Ala Val Val
                875                     880                     885
Asn His Gln Gly Thr Leu Glu Ser Gly His Tyr Thr Ser Phe Ile
                890                     895                     900
Arg His His Lys Asp Gln Trp Phe Lys Cys Asp Asp Ala Val Ile
                905                     910                     915
Thr Lys Ala Ser Ile Lys Asp Val Leu Asp Ser Glu Gly Tyr Leu
                920                     925                     930
Leu Phe Tyr His Lys Gln Val Leu Glu His Glu Ser Glu Lys Val
                935                     940                     945
Lys Glu Met Asn Thr Gln Ala Tyr
                950
```

<210> 22
<211> 2204
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 275791CB1

<400> 22

```
atatgccaat agacctgaca agtctgaatt ggaaactcag attgacagaa tgacgaagaa 60
gagctttagc agctgtcttg gagataagta agagagatgc ttcaccatct ctgagtcatg 120
aagatgatga taagccaact agcagcccag ataccggatt tgcagaagat gatattcaag 180
aaatgccgga aaatccagac actatggaaa ctgagaagcc caaaacaatc acagagctgg 240
atcctgccag ttttactgag ataactaaag actgtgatga gaataaagaa aacaaaactc 300
cagaaggatc tcagggagaa gttgattggc tccagcagta tgatatggag cgtgaaaggg 360
aagagcaaga gcttcagcag gcactggctc agagccttca agagcaagag gcttgggaac 420
agaaagaaga tgatgacctc aaaagagcta ccgagttaag tcttcaagag tttaacaact 480
cctttgtgga tgcattgggt tctgatgagg actctggaaa tgaggatgtt tttgatatgg 540
agtacacaga agctgaagct gaggaactga aaagaaatgc tgagacagga aatctgcctc 600
attcgtaccg gctcatcagt gttgtcagtc acattggtag cacttcttct tcaggtcatt 660
acattagtga tgtatatgac attaagaagc aagcgtggtt tacttacaat gacctggagg 720
tatcaaaaat ccaagaggct gccgtgcaga gtgatcgaga tcggagtggc tacatcttct 780
tttatatgca caaggagatc tttgatgagc tgctggaaac agaaaagaac tctcagtcac 840
ttagcacgga agtggggaag actacccgtc aggcctcgtg aggaacaaac tcctgggttg 900
gcagcatgca ctgcatattt gttactgctg cccacctcac ctttcctctg ctgaaggaga 960
atttggaatt ctacttgatg cgggagcaac aaacagctca gggccaaacc aaaagacaaa 1020
aattggagta acgtagaatg ctccatgcta ttttatggaa actttggtct cacatccgta 1080
gctgattatc ctctttttct cctatgagtg gcacttcttt tgtcttagga atacatgttg 1140
taaatatata tctgtgtatg tgtgtataca cacacacaga cacacacaca cacacacggg 1200
atgaatggag ccttaaagag ttaggatgag ccaccagaat atgcctgctc aaaattaata 1260
gcacagcagt ttggagaaga aatgaaggtg tcaaagagtc cattcacctg agaaatgtgt 1320
gaagacatac ttatcagttg gcttttagct tttatgttcc ttgagtagtt tcactcaagt 1380
ctgtaacctt ttgtgtttcc ttattagtaa aattcactgg aaagccagct cttcatgtta 1440
cactaatgac agtttgttct ctttgcaaga gaggggcatt actgtcacct gacttgagga 1500
gctgttttgt tgttgttgtt gtctgcaaat ttcatgaatt tgtgatgtct ttgctgttta 1560
catgcagtcc caagaaatgg attgttggtg ctttggaata tgttacagtc ccacatttga 1620
```

tatttcttat atactttgtt ttctctaagg agatttcttc acacagtatg ttcatcatat 1680
atcatcatca ttattatggt ggtaaagata gaatcttttt tcttttttttg tcattctggc 1740
catggagcag cattacccta atggattgca accaaaactt taaacaagta gaaagataat 1800
atttctccaa ttgggactcc ccagcaggaa tacttaggga taaggaagaa tgctagcatc 1860
tctgtctctc aaacataggg aggataagaa gagtgttctt ctggtaaagc taaaattctg 1920
gaccactgaa gctaaaagcc ctattgcaag tatgaaatta agtacttgag ctataggaca 1980
aaccttgggc atttaaccat ttactgtctg gctttgccct taaaataggg ttgcaattaa 2040
aatgtgattg gcttaggtaa tcccaaaaac taacaaataa caaaggtgca taatttattt 2100
atctactttt taggtgctct gagttgaggc aaagtagagc ggcaacatta agtgctatgc 2160
tagtcactta gctgacgtaa ccagctttgg taagcagctt atga 2204

<210> 23
<211> 2036
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1389845CB1

<400> 23
ccgatggggg ttaggctcca gggcttctgt cgagaccaag gatgcccaaa tatctgggtg 60
gtgggtgctg catacctggg ccctgggcag aacgaagggt atacagcctg ggccaccagg 120
ataagtccag aacccaccag gagctgagga cagacagaag gaccacggag ggggtgacgg 180
gctggtgtga ggattggtgc ccctgggcca ggactctcct ctcttctccc tgctggctcc 240
agaccagagt ccaagcccta ggcagtgcca cccttaccca gcccagcctt gaagacagaa 300
tgagaggggt ttcctgtctc caggtcctgc tccttctggt gctgggagct gctgggactc 360
agggaaggaa gtctgcagcc tgcgggcagc cccgcatgtc cagtcggatc gttgggggcc 420
gggatggccg ggacggagag tggccgtggc aggcgagcat ccagcatcgt ggggcacacg 480
tgtgcggggg gtcgctcatc gccccccagt gggtgctgac agcggcgcac tgcttcccca 540
ggagggcact gccagctgag taccgcgtgc gcctgggggc gctgcgtctg ggctccacct 600
cgccccgcac gctctcggtg cccgtgcgac gggtgctgct gcccccggac tactccgagg 660
acggggcccg cggcgacctg gcactgctgc agctgcgtcg cccggtgccc ctgagcgctc 720
gcgtccaacc cgtctgcctg cccgtgcccg gcgcccgccc gccgcccggc acaccatgcc 780
gggtcaccgg ctggggcagc ctccgcccag gagtgcccct cccagagtgg cgaccgctac 840
aaggagtaag ggtgccgctg ctggactcgc gcacctgcga cggcctctac cacgtgggcg 900
cggacgtgcc ccaggctgag cgcattgtgc tgcctgggag tctgtgtgcc ggctaccccc 960
agggccacaa ggacgcctgc caggtgtgca cccagcctcc ccagcctccg gagtcccctc 1020
cctgtgccca gcaccctccc tccctgaact ccaggaccca ggacatccca actcaggctc 1080
aggatcctgg cctccaacct agaggcacca cgccaggggt ctggaaccct gagaactgaa 1140
gtcctgggag ggctgggact taggctcctc tttctcctgc agggtgattc tgggggacct 1200
ctgacctgcc tgcagtctgg gagctgggtc ctggtgggcg tggtgagctg ggcaagggt 1260
tgtgccctgc ccaaccgtcc aggggtctac accagtgtgg ccacatatag cccctggatt 1320
caggctcgcg tcagcttcta atgctagccg gtgaggctga cctggagcca gctgctgggg 1380
tccctcagcc tcctggttca tccaggcacc tgcctatacc ccacatccct tctgcctcga 1440
ggccaagatg cctaaaaaag ctaaaggcca ccccaccccc cacccaccac ctcctgcctc 1500
ctctcctctt tggggatcac cagctctgac tccaccaacc ctcatccagg aatctgccat 1560
gagtcccagg gagtcacact ccccactccc ttcctggctt gtatttactt ttcttggccc 1620
tggccagggc tgggcgcaag gcacgcagtg atgggcaaac caattgctgc ccatctggcc 1680
tgtgtgccca tctttttctg gagaaagtca gattcacagc atgacagaga tttgacacca 1740
gggagatcct ccatagctgg ctttgaggac acggggacca cagccatgag cggcctctaa 1800
gagctgagag acagccggca gggaatcgga accctcagac ccacagccgc aaggcactgg 1860
attctggcag caccctgaag gagctgggaa gtaagttctt ccccagcctc cagataagag 1920
ccccgccggc caatcccttc atttcaacct aaagagaccc taagcagaga acctagctga 1980
gccactcctg acctacaaag ttgtgactta ataaatgtgt gctttaagct gctcca 2036

<210> 24
<211> 2185
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1726609CB1

<400> 24
gccatgcctc ctgcccacgg ccaccagcaa gctgtcgggc gcagtggagc agtggctgag 60

tgcagctgag cggctgtatg ggccctacat gtggggcagg tacgacattg tcttcctgcc 120
accctccttc cccatcgtgg ccatggagaa cccctgcctc accttcatca tctcctccat 180
cctggagagc gatgagttcc tggtcatcga tgtcatccac gaggtggccc acagttggtt 240
cggcaacgct gtcaccaacg ccacgtggga agagatgtgg ctgagcgagg gcctggccac 300
ctatgcccag cgccgtatca ccaccgagac ctacggtgct gccttcacct gcctggagac 360
tgccttccgc ctggacgccc tgcaccggca gatgaagctt ctgggagagg acagcccggt 420
cagcaaactg caggtcaagc tggagccagg agtgaatccc agccacctga tgaacctgtt 480
cacctacgag aagggctact gcttcgtgta ctacctgtcc cagctctgcg gagacccaca 540
gcgctttgat gactttctcc gagcctatgt ggagaagtac aagttcacca gcgtggtggc 600
ccaggacctg ctggactcct tcctgagctt cttcccggag ctgaaggagc agagcgtgga 660
ctgccgggca gggctggaat tcgagcgctg gctcaatgcc acaggcccgc cgctggctga 720
gccggacctg tctcagggat ccagcctgac ccggcccgtg gaggcccttt tccagctgtg 780
gaccgcagaa cctctggacc aggcagctgc ctcggccagc gccattgaca tctccaagtg 840
gaggaccttc cagacagcac tcttcctgga ccggctcctg gatgggtccc cgctgccgca 900
ggaggtggtg atgagcctgt ccaagtgcta ctcctccctg ctggactcga tgaacgctga 960
gatccgcatc cgctggctgc agattgtggt ccgcaacgac tactatcctg acctccacag 1020
ggtgcggcgc ttcctggaga gccagatgtc acgcatgtac accatcccgc tgtacgagga 1080
cctctgcacc ggtgccctca agtccttcgc gctggaggtc ttctaccaga cgcagggccg 1140
gctgcacccc aacctgcgca gagccatcca gcagatcctg tcccagggcc tgggctccag 1200
cacagagccc gcctcagagc ccagcacgga gctgggcaag gctgaagcag acacagactc 1260
ggacgcacag gccctgctgc ttggggacga ggcccccagc agtgccatct ctctcaggga 1320
cgtcaatgtg tctgcctagc cctgttggcg ggctgaccct cgacctccca gacaccacaa 1380
ttgtgccttc tgtgggccag gcctgccatg actgcgtctc ggctctggcc atgagctctg 1440
cccaggccca caagcccctc ccctgggctc tcccaggcag ggagaatggg gagagggacc 1500
tccttgtgtc tggcagagac ctgtggacct ggcctcccca ctcccagctc tcttgcactg 1560
caggccctgg ggccagcccg cacacaccat gcctcctgtc tcaacactga cagctgtgcc 1620
tagccccgga tgccagcacc tgccaggtgc cgccccgggg caagggcccc agcagcccta 1680
tggtgaccgc cacactgtgc cttaatgtct gccgggggcc caggctgtgc tgtccctgca 1740
gcacgcctcc ttgcagggat ctgagccacc ctccccgcac agccctgcac cccgccccta 1800
gggttggcag cctcagttgg cccctggcag aggaacaagg acacagacat tccctcagtg 1860
tggggggcag gggacacagg gagaggatgg ttgtccctgg ggagggccct ctggccccag 1920
gcaaccttag cccctcagaa cagggagtcc caggacccag ggagagtgtg gggacaggac 1980
agcctgtctc ttgtagcttc ctggggtggg aggcacaggg gcaaagcaat accccaggga 2040
aagtgggagg tggtgctggt gctctctcca ggcccaccat gctgggagag gcggccagag 2100
cctggggcct ccagcctggg actgctgtga tggggtatca cggtgatggt cccattaaac 2160
ttccactctg caaaaaaaaa aaaaa 2185

<210> 25
<211> 3486
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4503848CB1

<400> 25
ctgtcttaaa aaaagaggga gggaagatta ctgatttaat ttataaagga gaattattat 60
agctccaaca cctgacttta tttatctata tggtttaatt acacaaacaa ttcagtgttt 120
gaattataca aatttcatta aaactatgta attatgcaag aaaaatagga aatacagggg 180
cacttagttt tgtgcatatg tgttcacctg agagtatttg cttgtttttt taaaaaggtt 240
ctttttaatt taatatttaa ttttataatg cacattcata tgttgacttt ggaccaacag 300
aaatccctaa ttcttattct ttttctgatt ctttttagag ttggtggttc caggatttta 360
ctcagaatga cgttaggaag agaagtgatg tctcctcttc aggcaatgtc ttcctatact 420
gtggctggca gaaatgtttt aagatgggat ctttcaccag agcaaattaa aacaagaact 480
gaggagctca ttgtgcagac caaacaggtg tacgatgctg ttggaatgct cggtattgag 540
gaagtaactt acgagaactg tctgcaggca ctggcagatg tagaagtaaa gtatatagtg 600
gaaaggacca tgctagactt tccccagcat gtatcctctg acaaagaagt acgagcagca 660
agtacagaag cagacaaaag actttctcgt tttgatattg agatgagcat gagaggagat 720
atatttgaga gaattgttca tttacaggaa acctgtgatc tggggaagat aaaacctgag 780
gccagacgat acttggaaaa gtcaattaaa atggggaaaa gaaatgggct ccatcttcct 840
gaacaagtac agaatgaaat caaatcaatg aagaaaagaa tgagtgagct atgtattgat 900
tttaacaaaa acctcaatga ggatgatacc ttccttgtat tttccaaggc tgaacttggt 960
gctcttcctg atgatttcat tgacagttta gaaaagacag gtatgacaa gtataaaatt 1020
accttaaaat atccacacta tttccctgtc atgaagaaat gttgtatccc tgaaaccaga 1080
agaaggatgg aaatggcttt taatacaagg tgcaaagagg aaaacaccat aattttgcag 1140
cagctactcc cactgcgaac caaggtggcc aaactactcg gttatagcac acatgctgac 1200

ttcgtccttg aaatgaacac tgcaaagagc acaagccgcg taacagcctt tctagatgat 1260
ttaagccaga agttaaaacc cttgggtgaa gcagaacgag agtttatttt gaatttgaag 1320
aaaaaggaat gcaaagacag ggttttgaa tatgatggga aaatcaatgc ctgggatcta 1380
tattactaca tgactcagac agaggaactc aagtattcca tagaccaaga gttcctcaag 1440
gaatacttcc caattgaggt ggtcactgaa ggcttgctga acacctacca ggagttgttg 1500
ggactttcat ttgaacaaat gacagatgct catgtttgga acaagagtgt tacactttat 1560
actgtgaagg ataaagctac aggagaagta ttgggacagt tctatttgga cctctatcca 1620
agggaaggaa aatacaatca tgcggcctgc ttcggtctcc agcctggctg ccttctgcct 1680
gatggaagcc ggatgatggc agtggctgcc ctcgtggtga acttctcaca gccagtggca 1740
ggtcgtccct ctctcctgag acacgacgag gtgaggactt actttcatga gtttggtcac 1800
gtgatgcatc agatttgtgc acagactgat tttgcacgat ttagcggaac aaatgtggaa 1860
actgactttg tagaggtgcc atcgcaaatg cttgaaaatt gggtgtggga cgtcgattcc 1920
ctccgaagat tgtcaaaaca ttataaagat ggaagcccta ttgcagacga tctgcttgaa 1980
aaacttgttg cttctaggct ggtcaacaca ggtcttctga ccctgcgcca gattgttttg 2040
agcaaagttg atcagtctct tcataccaac acatcgctgg atgctgcaag tgaatatgcc 2100
aaatactgct cagaaatatt aggagttgca gctactccag gcacaaatat gccagctacc 2160
tttggacatt tggcaggggg atacgatggc caatattatg gatatctttg gagtgaagta 2220
ttttccatgg atatgtttta cagctgtttt aaaaaagaag ggataatgaa tccggaggtt 2280
ggaatgaaat acagaaacct aatcctgaaa cctgggggat ctctggacgg catggacatg 2340
ctccacaatt tcttgaaacg tgagccaaac caaaaagcgt tcctaatgag tagaggcctg 2400
catgctccgt gaactgggga tctttggtag ccgtccatgt ctggaggaca agtcgacatc 2460
accatgtgtt actggcctgg aaactgaagg gagttttgca agtgaaaatt tagatttcta 2520
ttgacatcct tttgttttct aattttaaaa attataaaga tgtaaatgga attataaata 2580
ctgtgaccta agaaaagacc cactagaaag taattgtact ataaaatttc ataaaactgg 2640
atttgatttc tttttatgaa agtttcatat gaatgtaact tgattttta ctattataat 2700
ctagataata tgatataaga gggctaagaa tttttaaatt gaatcatata tatgatataa 2760
tttgatcctt cttgtatctt gaagttttgt acttgggatt tctggactga taaatgaatc 2820
atcacattct tctggtaaat attttcttgg agctctgtgt caactttgat cctttgtctc 2880
ccaggaaggt gtgacctctc ctttgcctgc atacctcaag gccaggggaa tatgcctcag 2940
tgatgcattt atctttgtat atcaggccgc atgattccca actttctgcc acacttaaat 3000
tacgttcctc catttcagtt ttgtcttttc tgtctaaagt tcagtcaaag agtatcaaaa 3060
aattatgttt cagctagact ggtgtaatgt ataagttttt gtatcttgta ttagaggatt 3120
tcgtagcttt tattagaggc tcatttccac ctcagcatac aagatcgtta gtcttttggc 3180
atgtgtgcca attagaatac taaagcaagt ccaagcacat ttttctcttc tcacgtttct 3240
aataagtgtt agggactttg cctcttttac ttaccacgtc cccaaaagtg tcaggtagac 3300
atgtcacaaa tggctctgta gagagccatg ggaagagaga ggaggtggat gtggaacata 3360
aagggttcag aaactccaga agaggagtgg gttttggata gaagcatttg aggacagctg 3420
ctccaaagcc ttatgtgtat gatgaaactt aaccacgggg aagagactct tcagtagcct 3480
gttctg 3486

<210> 26
<211> 2847
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<223> Incyte ID No: 5544089CB1

<400> 26
caatgacgct tggacgagga tttatttcta caagctaatt gaatccagga gcagctttaa 60
ttattaacac taacggaaga gaaaaggagt atttccaagg gctcaaatgg aagctgtact 120
cagtccggtg gaggcagggg gaggtaaagt ttctcacact caagtcgtct tcatagttta 180
ctgtcctttt ccaaacaaaa gctaataacg ccatacgcat ccacacactc cctcctggat 240
gaacctaagt ctcgtcccca ctgtcacccc aaggccagtt atcaaaaact gttccttctc 300
tgccctcaaa gactgaagcc gcaggccctg ttctgcctct gctcaggaat ctgattgctc 360
ttaaagtgct cttacaagat tccgtcgatg tttgctccct ctgtcttgtc atcaggacta 420
agtggtggag catcaaaagg tagaaagatg gaacttattc agccaaagga gccaacttca 480
cagtacattt ctctttgtca tgaattgcat actttgttcc aagtcatgtg gtctggaaag 540
tgggcgttgg tctcaccatt tgctatgcta cactcagtgt ggagactcat tcctgccttt 600
cgtggttacg cccaacaaga cgctcaggaa tttctttgtg aactttaga taaaatacaa 660
cgtgaattag agacaactgg taccagttta ccagctctta tccccacttc tcaaaggaaa 720
ctcatcaaac aagttctgaa tgttgtaaat aacatttttc atggacaact tcttagtcag 780
gttacatgtc ttgcatgtga caacaaatca aataccatag aacctttctg ggacttgtca 840
ttggagtttc cagaaaggta tcaatgcagt ggaaaagata ttgcttccca gccatgtctg 900
gttactgaaa tgttggccaa atttacagaa actgaagctt tagaaggaaa aatctacgta 960
tgtgaccagt gtaactcaaa gcgtagaagg ttttcctcca aaccagttgt actcacagaa 1020

gcccagaaac aacttatgat atgccaccta cctcaggttc tcagactgca cctcaaacga 1080
ttcaggtggt caggacgtaa taaccgagag aagattggtg ttcatgttgg ctttgaggaa 1140
atcttaaaca tggagcccta ttgctgcagg gagaccctga aatccctcag accagaatgc 1200
tttatctatg acttgtccgc ggtggtgatg caccatggga aaggatttgg ctcagggcac 1260
tacactgcct actgctataa ttctgaagga gggttctggg tacactgcaa tgattccaaa 1320
ctaagcatgt gcactatgga tgaagtatgc aaggctcaag cttatatctt gttttatacc 1380
caacgagtta ctgagaatgg acattctaaa cttttgcctc cagagctcct gttggggagc 1440
caacatccca atgaagacgc tgatacctcg tctaatgaaa tccttagctg atccaaagac 1500
aatggggttt tcttcctgtg atttatatat atacttttta aaagactgat gtaccatttt 1560
aaacttcatt ttttcttgtg aatcagtgta tactacattt atacatttta tatctaacaa 1620
tttttttttt tacaaagtat aaatgtatat atcaactgaa ggtaactact tttttcatat 1680
ttggagtttt aaacttttgg tgtttacctc agactgatgt tacctctttt atatttttat 1740
gtcttaattg gctcggatga tgaacttgtg caatcttcta ccaacaaagt tcaagtggca 1800
tcattttata tacatgtatc tttttcaggt atttctata caaattctta atagatggaa 1860
aattagactc tactttggtc actaatagtc tttcatttgt atattgaagt taccttgccc 1920
cttggagtta ttgaagtgac atgtcaaggt atcacctaaa tattcttcag tcacactcac 1980
tggtatttct gaggctttgt gtgttaacag gccttgtaat tgacattatt ttggttaatg 2040
taaccccaaa attgctttag taattgctct ttggcatagt caaactataa atgaaaatgg 2100
cagctttaca aatagtatat ttaagtgaac tctggaacta tggacatgaa aaaaatgatg 2160
gctgggattt atgatttttg tctggcagca aacaggtttg tccagaagtc taataattaa 2220
gcagtcataa aaagtctgaa tttagtaaac cagtgtatga tgttattcaa atagtttacc 2280
ttgggtatga gttcatttta taatgtctga tgacattaga tctcttaaaa ctttatgtat 2340
ttttttagt tcaaaggaat agagtcttga agagaaaaaa ttatagggca gaaaagataa 2400
gtgttcaaaa ttggcaactg gactattatt atgtctagca tctcattcta aataactaaa 2460
gcttgattta ctcttgctag gattatgtga ctactaggta ggagcctctt aaaacactgg 2520
ccctgagcat taaaaaaaaa aaaaaaaact aaaagctatc tatctaaact tgcaaaaaaa 2580
aaattccggt gggggtcacc ctttccttc ttctgaaaat ctcacggggt ttctttaaag 2640
ccctgttgct gcaaacttta tcttttttgg gggggtaga atcacctaat ctctgtagac 2700
cagctatgtt tctaagctct gttaaccacg gggagatctg gtacccttt tttaaaaggg 2760
ggtttatttg cgggttgaag tcttagtgaa aagtagtccc ctggagaatg cggtccaccc 2820
ctgggggcca tctgttaggt aaaactt 2847

<210> 27
<211> 890
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474081CB1

<400> 27
gaggccaaga attcggcacg aggcacttac tccctgagct aagggggaag agctggatca 60
ccatgaaata tgtcttctat ttgggtgtcc tcgctgggac atttttcttt gctgactcat 120
ctgttcagaa agaagaccct gctccctatt tggtgtacct caagtctcac ttcaacccct 180
gtgtgggcgt cctcatcaaa cccagctggg tgctggcccc agctcactgc tatttaccaa 240
atctgaaagt gatgctggga aatttcaaga gcagagtcag agacggtact gaacagacaa 300
ttaaccccat tcagatcgtc cgctactgga actacagtca tagcgcccca caggatgacc 360
tcatgctcat caagctggct aagcctgcca tgctcaatcc caaagtccag ccccttaccc 420
tcgccaccac caatgtcagg ccaggcactg tctgtctact ctcaggtttg gactggagcc 480
aagaaaacag tggccgacac cctgacttgc ggcagaacct ggaggccccc gtgatgtctg 540
atcgagaatg ccaaaaaaca gaacaaggaa aaagccacag gaattcctta tgtgtgaaat 600
ttgtgaaagt attcagccga atttttgggg aggtggccgt tgctactgtc atctgcaaag 660
acaagctcca gggaatcgag gtggggcact tcatgggagg ggacgtcggc atctacacca 720
atgtttacaa atatgtatcc tggattgaga acactgctaa ggacaagtga gaccctactt 780
ctccctctgc attccactgg ctctgccatg gactatacaa gcagataatt ttccctctat 840
tcaaaataaa atctccaaat gaaaatttgg gaatgtagca aaaaaaaaa 890

<210> 28
<211> 1577
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5281209CB1

<400> 28
atgcagccca cgggccgcga gggttcccgc gcgctcagcc ggcggtatct gcggcgtctg 60
ctgctcctgc tactgctgct gctgctgcgg cagcccgtaa cccgcgcgga gaccacgccg 120
ggcgccccca gagccctctc cacgctgggc tcccccagcc tcttcaccac gccgggtgtc 180
cccagcgccc tcactacccc aggcctcact acgccaggca cccccaaaac cctggacctt 240
cggggtcgcg cgcaggccct gatgcggagt ttcccactcg tggacggcca caatgacctg 300
ccccaggtcc tgagacagcg ttacaagaat gtgcttcagg atgttaacct gcgaaatttc 360
agccatggtc agaccagcct ggacaggctt agagacggcc tcgtgggtgc ccagttctgg 420
tcagcctccg tctcatgcca gtcccaggac cagactgccg tgcgcctcgc cctggagcag 480
attgacctca ttcaccgcat gtgtgcctcc tactctgaac tcgagcttgt gacctcagct 540
gaaggtctga acagctctca aaagctggcc tgcctcattg gcgtggaggg tggtcactca 600
ctggacagca gcctctctgt gctgcgcagt ttctatgtgc tgggggtgcg ctacctgaca 660
cttaccttca cctgcagtac accatgggca gagagttcca ccaagttcag acaccacatg 720
tacaccaacg tcagcggatt gacaagcttt ggtgagaaag tagtagagga gttgaaccgc 780
ctgggcatga tgatagattt gtcctatgca tcggacacct tgataagaag ggtcctggaa 840
gtgtctcagg ctcctgtgat cttctcccac tcagctgcca gagctgtgtg tgacaatttg 900
ttgaatgttc ccgatgatat cctgcagctt ctgaagaaga acggtggcat cgtgatggtg 960
acactgtcca tgggggtgct gcagtgcaac ctgcttgcta acgtgtccac tgtggcagat 1020
cactttgacc acatcagggc agtcattgga tctgagttca tcgggattgg tggaaattat 1080
gacgggactg gccggttccc tcaggggctg gaggatgtgt ccacataccc agtcctgata 1140
gaggagttgc tgagtcgtag ctggagcgag gaagagcttc aaggtgtcct tcgtggaaac 1200
ctgctgcggg tcttcagaca agtggaaaag gtgagagagg agagcagggc gcagagcccc 1260
gtggaggctg agtttccata tgggcaactg agcacatcct gccactccca cctcgtgcct 1320
cagaatggac accaggctac tcatctggag gtgaccaagc agccaaccaa tcgggtcccc 1380
tggaggtcct caaatgcctc cccatacctt gttccaggcc ttgtggctgc tgccaccatc 1440
ccaaccttca cccagtggct ctgctgacac agtcggtccc cgcagaggtc actgtggcaa 1500
agcctcacaa agccccctct cctagttcat tcacaagcat atgctgagaa taaacatgtt 1560
acacatggaa aaaaaaa 1577

<210> 29
<211> 1958
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2256251CB1

<400> 29
aagcggtcga gctcggcatt cattgtaacg gcgccatgtg ctggaaaggt cgtgtggttt 60
ctgctcgcat ctctcggttg agtggggctg gtcggggtgt gctgcagggc tgtctccccc 120
accaccactg tagtcagtct gtaccttggg agatgctctg aggccatgaa acaacctggc 180
cctcctcgaa ctttctcccc acagcgtccc caccgtggcc ctggaaccag ctgggggctt 240
tgccgtgtgg gagagccggt gccccagccc acacccgctg cctatctata aacatctctg 300
tctgtctaca tcccagcttc ccttccattc agcccagtgg gcacactcca tcaccagcac 360
aattatccag ctcaggcaga cccaggtgtg gccagtgggt ctccagctga cttcctctta 420
atttcttctt aacttactga ggtgaagttt acagaagata aagttacatt atcaagcgtc 480
caattcagag gccccgagca ccttggacag tgctgtctgc cccccgcccc ccgaatttta 540
tccagcttca aatatctcca ccacccctga aggaaaaccg gggcccacta ggcagccacc 600
cctagcaccc cggccttct cgggggctcc accgttctga gcccctttct agccgcctag 660
gggccctctg cagcctttcc accgcctccg ggagccctgg ttagtttgtg gagcatggtg 720
cttagaaaag acgacctgag ccccaggcgc gctcactgct cctgagacgt catttctgct 780
gcacccacga tgcttctggg gagagtctgg cagacgagag agctgaagag caaagtcccc 840
aagaaggcag ggaggtgtgg tcagggaagg cttcatggag gaagtgcagt gggcttcttg 900
ggatccccac caggcacccc ttcctccttc gacttagggt gtggccggcc gcaggtttcg 960
gatgcaggcg gccggatcgt gggggtcac gctgccccgg ccggcgcatg gccatggcag 1020
gccagcctcc gcctgcggag ggtgcacgtg tgcggcgggt cactgctcag cccccagtgg 1080
gtgctcacag ctgcccactg cttctccggg tccctgaact catccgacta ccaggtgcac 1140
ctgggggaac tggagatcac tctgtctccc cacttctcca ccgtgaggca gatcatcctg 1200
cactccagcc cctcaggaca gccggggacc agcggggaca tcgccctggt ggagctcagt 1260
gtccccgtga ccctcttcag ccggatcctg cccgtctgcc tcccggaggc ctcagatgac 1320
ttctgccctg ggatccggtg ctgggtgacc ggctggggct atacgcggga gggagagcct 1380
ctgccacccc cgtacagcct gcgggaggtg aaagtctccg tggtggacac agagacctgc 1440
cgccgggact atcccggccc cgggggcagc atccttcagc ccgacatgct gtgtgcccgg 1500
ggccccgggg atgcctgcca ggacgactcc gggggcctc tggtctgcca ggtgaacggt 1560
gcctgggtgc aggctggcat tgtgagctgg ggtgagggct gcggccgccc caacaggccg 1620
ggagtctaca ctcgtgtccc tgcctacgtg aactggatcc gccgccacat cacagcatca 1680

gggggctcag agtctgggta ccccaggctc cccctcctgg ctggcttatt cctccccggc 1740
ctcttccttc tgctagtctc ctgtgtcctg ctggccaagt gcctgctgca cccatctgcg 1800
gatggtactc ccttccccgc ccctgactga tggcaggaat ccaagtgcat ttcttaaata 1860
agttactatt tattccgctc cgccccctcc ctctcccttg agaagctgag tcttctgcat 1920
cagattattg caacatttaa cctgaattta acgacacc 1958

<210> 30
<211> 3106
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7160544CB1

<400> 30
gctccgaggc caaggccgct gctactgccg ccgctgcttc ttagtgccgc gttcgccgcc 60
tgggttgtca ccggcgccgc cgctgaggaa gccactgcaa ccaggaccgg agtggaggcg 120
gcgcagcatg aagcggcgca ggcccgctcc atagcgcacg tcgggacggt ccgggcgggg 180
ccggggggaa ggaaaatgca acatggcagc agcaatggaa acagaacagc tgggtgttga 240
gatatttgaa actgcggact gtgaggagaa tattgaatca caggatcggc ctaaattgga 300
gccttttat gttgagcggt attcctggag tcagcttaaa aagctgcttg ccgataccag 360
aaaatatcat ggctacatga tggctaaggc accacatgat ttcatgtttg tgaagaggaa 420
tgatccagat ggacctcatt cagacagaat ctattacctt gccatgtctg gtgagaacag 480
agaaaataca ctgttttatt ctgaaattcc caaaactatc aatagagcag cagtcttaat 540
gctctcttgg aagcctcttt tggatctttt tcaggcaaca ctggactatg gaatgtattc 600
tcgagaagaa gaactattaa gagaaagaaa acgcattgga acagtcggaa ttgcttctta 660
cgattatcac caaggaagtg gaacatttct gtttcaagcc ggtagtggaa tttatcacgt 720
aaaagatgga gggccacaag gatttacgca acaaccttta aggcccaatc tagtggaaac 780
tagttgtccc aacatacgga tggatccaaa attatgccct gctgatccag actggattgc 840
ttttatacat agcaacgata tttggatatc taacatcgta accagagaag aaaggagact 900
cacttatgtg cacaatgagc tagccaacat ggaagaagat gccagatcag ctggagtcgc 960
tacctttgtt ctccaagaag aatttgatag atattctggc tattggtggt gtccaaaagc 1020
tgaaacaact cccagtggtg gtaaaattct tagaattcta tatgaagaaa atgatgaatc 1080
tgaggtggaa attattcatg ttacatcccc tatgttggaa acaaggaggg cagattcatt 1140
ccgttatcct aaaacaggta cagcaaatcc taaagtcact tttaagatgt cagaaataat 1200
gattgatgct gaaggaagga tcatagatgt catagataag gaactaattc aacctttga 1260
gattctattt gaaggagttg aatatattgc cagagctgga tggactcctg agggaaaata 1320
tgcttggtcc atcctactag atcgctccca gactcgccta cagatagtgt tgatctcacc 1380
tgaattattt atcccagtag aagatgatgt tatggaaagg cagagactca ttgagtcagt 1440
gcctgattct gtgacgccac taattatcta tgaagaaaca acagacatct ggataaatat 1500
ccatgacatc tttcatgttt ttccccaaag tcacgaagag gaaattgagt ttatttttgc 1560
ctctgaatgc aaaacaggtt tccgtcattt atacaaaatt acatctattt taaaggaaag 1620
caaatataaa cgatccagtg gtgggctgcc tgctccaagt gatttcaagt gtcctatcaa 1680
agaggagata gcaattacca gtggtgaatg ggaagttctt ggccggcatg gatctaatat 1740
ccaagttgat gaagtcagaa ggctggtata ttttgaaggc accaaagact cccctttaga 1800
gcatcacctg tacgtagtca gttacgtaaa tcctggagag gtgacaaggc tgactgaccg 1860
tggctactca cattcttgct gcatcagtca gcactgtgac ttctttataa gtaagtatag 1920
taaccagaag aatccacact gtgtgtccct ttacaagcta tcaagtcctg aagatgaccc 1980
aacttgcaaa acaaaggaat tttgggccac cattttggat tcagcaggtc ctcttcctga 2040
ctatactcct ccagaaattt tctcttttga aagtactact ggatttacat tgtatgggat 2100
gctctacaag cctcatgatc tacagcctgg aaagaaatat cctactgtgc tgttcatata 2160
tggtggtcct caggtgcagt tggtgaataa tcggtttaaa ggagtcaagt atttccgctt 2220
gaataccctа gcctctctag gttatgtggt tgtagtgata gacaacaggg gatcctgtca 2280
ccgagggctt aaatttgaag gcgcctttaa atataaaatg ggtcaaatag aaattgacga 2340
tcaggtggaa ggactccaat atctagcttc tcgatatgat ttcattgact tagatcgtgt 2400
gggcatccac ggctggtcct atggaggata cctctccctg atggcattaa tgcagaggtc 2460
agatatcttc agggttgcta ttgctggggc cccagtcact ctgtggatct tctatgatac 2520
aggatacacg gaacgttata tgggtcaccc tgaccagaat gaacagggct attacttagg 2580
atctgtggcc atgcaagcag aaaagttccc ctctgaacca aatcgtttac tgctcttaca 2640
tggtttcctg gatgagaatg tccattttgc acataccagt atattactga gtttttttagt 2700
gagggctgga aagccatatg atttacagat ctatcctcag gagagacaca gcataagagt 2760
tcctgaatcg ggagaacatt atgaactgca tcttttgcac taccttcaag aaaaccttgg 2820
atcacgtatt gctgctctaa aagtgatata attttgacct gtgtagaact ctctggtata 2880
cactggctat ttaaccaaat gaggaggttt aatcaacaga aaacacagaa ttgatcatca 2940
cattttgata cctgccatgt aacatctact cctgaaaata aatgtggtgc catgcagggg 3000
tctacggttt gtggtagtaa tctaatacct taaccccaca tgctcaaaat caaatgatac 3060

atattcctga gagacccagc aataccataa gaattactaa aaaaaa 3106

<210> 31
<211> 3567
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477386CB1

<400> 31
atggctccac tccgcgcgct gctgtcctac ctgctgcctt tgcactgtgc gctctgcgcc 60
gccgcgggca gccggacccc agagctgcac ctctctggaa agctcagtga ctatggtgtg 120
acagtgccct gcagcacaga ctttcgggga cgcttcctct cccacgtggt gtctggccca 180
gcagcagcct ctgcagggag catggtagtg gacacgccac ccacactacc acgacactcc 240
agtcacctcc gggtggctcg cagccctctg cacccaggag ggaccctgtg gcctggcagg 300
gtggggcgcc actccctcta cttcaatgtc actgttttcg ggaaggaact gcacttgcgc 360
ctgcggccca atcggaggtt ggtagtgcca ggatcctcag tggagtggca ggaggatttt 420
cgggagctgt tccggcagcc cttacggcag gagtgtgtgt acactggagg tgtcactgga 480
atgcctgggg cagctgttgc catcagcaac tgtgacggat tggcgggcct catccgcaca 540
gacagcaccg acttcttcat tgagcctctg gagcggggcc agcaggagaa ggaggccagc 600
gggaggacac atgtggtgta ccgccgggag gccgtccagc aggagtgggc agaacctgac 660
ggggacctgc acaatgaagc ctttggcctg ggagaccttc ccaacctgct gggcctggtg 720
ggggaccagc tgggcgacac agagcggaag cggcggcatg ccaagccagg cagctacagc 780
atcgaggtgc tgctggtggt ggacgactcg gtggttcgct tccatggcaa ggagcatgtg 840
cagaactatg tcctcaccct catgaatatc gtggtagatg agatttacca cgatgagtcc 900
ctgggggttc atataaatat tgccctcgtc cgcttgatca tggttggcta ccgacagcag 960
tccctgagcc tgatcgagcg cgggaacccc tcacgcagcc tggagcaggt gtgtcgctgg 1020
gcacactccc agcagcgcca ggaccccagc cacgctgagc accatgacca cgttgtgttc 1080
ctcacccggc aggactttgg gccctcaggt gggtatgcac ccgtcactgg catgtgtcac 1140
cccctgagga gctgtgccct caaccatgag gatggcttct cctcagcctt cgtgatagct 1200
catgagaccg gccacgtgct cggcatggag catgacggtc aggggaatgg ctgtgcagat 1260
gagaccagcc tgggcagcgt catggcgccc ctggtgcagg ctgccttcca ccgcttccat 1320
tggtcccgct gcagcaagct ggagctcagc cgctacctcc cgtcctacga ctgcctcctc 1380
gatgacccct ttgatcctgc ctggccccag cccccagagc tgcctgggat caactactca 1440
atggatgagc agtgccgctt tgactttggc agtggctacc agacctgctt ggcattcagg 1500
acctttgagc cctgcaagca gctgtggtgc agccatcctg acaacccgta cttctgcaag 1560
accaagaagg ggcccccgct ggatgggact gagtgtgcac ccggcaagtg gtgcttcaaa 1620
ggtcactgca tctggaagtc gccggagcag acatatggcc aggatggagg ctggagctcc 1680
tggaccaagt ttgggtcatg ttcgcggtca tgtgggggcg gggtgcgatc ccgcagccgg 1740
agctgcaaca acccctcgcc agcctatgga ggccgcctgt gcttagggcc catgttcgag 1800
taccaggtct gcaacagcga ggagtgccct gggacctacg aggacttccg ggcccagcag 1860
tgtgccaagc gcaactccta ctatgtgcac cagaatgcca agcacagctg ggtgccctac 1920
gagcctgacg atgacgccca gaagtgtgag ctgatctgcc agtcggcgga cacgggggac 1980
gtggtgttca tgaaccaggt ggttcacgat gggacacgct gcagctaccg ggacccatac 2040
agcgtctgtg cgcgtggcga gtgtgtgcct gtcggctgtg acaaggaggt ggggtccatg 2100
aaggcggatg acaagtgtgg agtctgcggg ggtgacaact cccactgcag gactgtgaag 2160
gggacgctgg gcaaggcctc caagcaggca ggtgctctca agctggtgca gatcccagca 2220
ggtgccaggc acatccagat tgaggcactg gagaagtccc cccaccgcat tgtggtgaag 2280
aaccaggtca ccggcagctt catcctcaac cccaagggca aggaagccac aagccggacc 2340
ttcaccgcca tgggcctgga gtgggaggat gcggtggagg atgccaagga aagcctcaag 2400
accagcgggc ccctgcctga agccattgcc atcctggctc tcccccaac tgagggtggc 2460
ccccgcagca gcctggccta caagtacgtc atccatgagg acctgctgcc ccttatcggg 2520
agcaacaatg tgctcctgga ggagatggac acctatgagt gggcgctcaa gagctgggcc 2580
ccctgcagca aggcctgtgg aggaggtatc cagttcacca aatacggctg ccggcgcaga 2640
cgagaccacc acatggtgca gcgacacctg tgtgaccaca agaagaggcc caagcccatc 2700
cgccggcgct gcaaccagca cccgtgctct cagcctgtgt gggtgacgga ggagtggggt 2760
gcctgcagcc ggagctgtgg gaagctgggg gtgcagacac gggggataca gtgcctgctg 2820
cccctctcca atggaaccca caaggtcatg ccggccaaag cctgcgccgg ggaccggcct 2880
gaggcccgac ggccctgtct ccgagtgccc tgcccagccc agtggaggct gggagcctgg 2940
tcccagtgct ctgccacctg tggagagggc atccagcagc ggcaggtggt gtgcaggacc 3000
aacgccaaca gcctcgggca ttgcgagggg gataggccag acactgtcca ggtctgcagc 3060
ctgcccgcct gtggagcgga gccctgcacg ggagacaggt ctgtcttctg ccagatggaa 3120
gtgctcgatc gctactgctc cattcccggc taccaccggc tctgctgtgt gtcctgcatc 3180
aagaaggcct cgggccccaa ccctggccca gaccctggcc caacctcact gcccccccttc 3240
tccactcctg gaagcccctt accaggaccc caggaccctg cagatgctgc agagcctcct 3300

ggaaagccaa cgggatcaga ggaccatcag catggccgag ccacacagct cccaggagct 3360
ctggatacaa gctccccagg gacccagcat ccctttgccc ctgagacacc aatccctgga 3420
gcatcctgga gcatctcccc taccacccc gggggctgc cttggggctg gactcagaca 3480
cctacgccag tccctgagga caaagggcaa cctggagaag acctgagaca tcccggcacc 3540
agcctccctg ctgcctcccc ggtgaca 3567

<210> 32
<211> 2930
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473089CB1

<400> 32
cacgcagacc gcggcagcgg ccgagagccc ggcccagccc cttcccacag cgcggcgttg 60
cgctgccccgg cgccatgctt ctgctgggca tcctaaccct ggctttcgcc gggcgaaccg 120
ctggaggctc tgagccagag cgggaggtag tcgttcccat ccgactggac ccggacatta 180
acggccgccg ctactactgg cggggtcccg aggactccgg ggatcaggga ctcatttttc 240
agatcacagc atttcaggag gacttttacc tacacctgac gccggatgct cagttcttgg 300
ctcccgcctt ctccactgag catctgggcg tcccccctcca ggggctcacc gggggctctt 360
cagacctgcg acgctgcttc tattctgggg acgtgaacgc cgagccggac tcgttcgctg 420
ctgtgagcct gtgcgggggg ctccgcggag cctttggcta ccgaggcgcc gagtatgtca 480
ttagcccgct gcccaatgct agcgcgccgg cggcgcagcg caacagccag ggcgcacacc 540
ttctccagcg ccggggtgtt ccgggcgggc cttccggaga ccccacctct cgctgcgggg 600
tggcctcggg ctggaacccc gccatcctac gggccctgga cccttacaag ccgcggcggg 660
cgggcttcgg ggagagtcgt agccggcgca ggtctgggcg cgccaagcgt ttcgtgtcta 720
tcccgcggta cgtggagacg ctggtggtcg cggacgagtc aatggtcaag ttccacggcg 780
cggacctgga acattatctg ctgacgctgc tggcaacggc ggcgcgactc taccgccatc 840
ccagcatcct caaccccatc aacatcgttg tggtcaaggt gctgcttctt agagatcgtg 900
actccgggcc caaggtcacc ggcaatgcgg ccctgacgct gcgcaacttc tgtgcctggc 960
agaagaagct gaacaaagtg agtgacaagc accccgagta ctgggacact gccatcctct 1020
tcaccaggca ggacctgtgt ggagccacca cctgtgacac cctgggcatg gctgatgtgg 1080
gtaccatgtg tgaccccaag agaagctgct ctgtcattga ggacgatggg cttccatcag 1140
ccttcaccac tgcccacgag ctgggccacg tgttcaacat gccccatgac aatgtgaaag 1200
tctgtgagga ggtgtttggg aagctccgag ccaaccacat gatgtccccg accctcatcc 1260
agatcgaccg tgccaacccc tggtcagcct gcagtgctgc catcatcacc gacttcctgg 1320
acagcgggca cggtgactgc ctcctggacc aacccagcaa gcccatctcc ctgcccgagg 1380
atctgccggg cgccagctac accctgagcc agcagtgcga gctggctttt ggcgtgggct 1440
ccaagccctg tccttacatg cagtactgca ccaagctgtg gtgcaccggg aaggccaagg 1500
gacagatggt gtgccagacc cgccacttcc cctgggccga tggcaccagc tgtggcgagg 1560
gcaagctctg cctcaaaggg gcctgcgtgg agagacacaa cctcaacaag cacagggtgg 1620
atggttcctg ggccaaatgg gatccctatg gcccctgctc gcgcacatgt ggtgggggcg 1680
tgcagctggc caggaggcag tgcaccaacc ccacccctgc caacgggggc aagtactgcg 1740
agggagtgag ggtgaaatac cgatcctgca atctggagcc ctgccccagc tcagcctccg 1800
gaaagagctt ccgggaggag cagtgtgagg ctttcaacgg ctacaaccac agcaccaacc 1860
ggctcactct cgccgtggca tgggtgccca agtactccgg cgtgtctccc cgggacaagt 1920
gcaagctcat ctgccgagcc aatggcactg gctacttcta tgtgctggca cccaaggtgg 1980
tggtggacgg cacgctgtgc tctcctgact ccacctccgt ctgtgtccaa ggcaagtgca 2040
tcaaggctgg ctgtgatggg aacctgggct ccaagaagag attcgacaag tgtggggtgt 2100
gtgggggaga caataagagc tgcaagaagg tgactggact cttcaccaag cccatgcatg 2160
gctacaattt cgtggtggcc atccccgcag gcgcctcaag catcgacatc cgccagcgcg 2220
gttacaaagg gctgatcggg gatgacaact acctggctct gaagaacagc caaggcaagt 2280
acctgctcaa cgggcatttc gtggtgtcgg cggtggagcg ggacctggtg gtgaagggca 2340
gtctgctgcg gtacagcggc acgggcacag cggtggagag cctgcaggct tcccggccca 2400
tcctggagcc gctgaccgtg gaggtcctct ccgtggggaa gatgacaccg ccccgggtcc 2460
gctactcctt ctatctgccc aaagagcctc gggaggacaa gtcctctcat cccccgcacc 2520
cccggggagg accctctgtc ttgcacaaca gcgtcctcag cctctccaac caggtggagc 2580
agccggacga caggcccccct gcacgctggg tggctggcag ctgggggccg tgctccgcga 2640
gctgcggcag tggcctgcag aagcgggcgg tggactggcg gggctccgcc gggcagcgca 2700
cggtccctgc ctgtgatgca gcccatcggc ccgtggagac acaagcctgc ggggagccct 2760
gccccacctg ggagctcagc gcctggtcac cctgctccaa gagctgcggc cggggatttc 2820
agaggcgctc actcaagtgt gtgggccacg gaggccggct gctggcccgg gaccagtgca 2880
acttgcaccg caagccccag gagctggact tctgcgtcct gaggccgtgc 2930

<210> 33

<211> 4230
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7604035CB1

<400> 33
agcgaggttg cctggagaga gcgcctgggc gcagaagggt taacgggcca ccgggggctc 60
gcagagcagg agggtgctct cggacggtgt gtccccсact gcactcctga acttggagga 120
cagggtcgcc gcgagggacg cagagagcac cctccacgcc cagatgcctg cgtagttttt 180
gtgaccagtc cgctcctgcc tccccctggg gcagtagagg gggagcgatg gagaactgga 240
ctggcaggcc ctggctgtat ctgctgctgc ttctgtccct ccctcagctc tgcttggatc 300
aggaggtgtt gtccggacac tctcttcaga cacctacaga ggagggccag ggccccgaag 360
gtgtctgggg accttgggtc cagtgggcct cttgctccca gccctgcggg gtgggggtgc 420
agcgcaggag ccggacatgt cagctcccta cagtgcagct ccacccgagt ctgcccctcc 480
ctccccggcc cccaagacat ccagaagccc tcctccccсg gggccagggt cccagacccc 540
agacttctcc agaaaccctc cccttgtaca ggacacagtc tcggggaagg ggtggcccac 600
ttcgaggtcc cgcttcccac ctagggagag aggagaccca ggagattcga gcggccagga 660
ggtcccggct tcgagacccc atcaagccag gaatgttcgg ttatgggaga gtgccctttg 720
cattgccact gcaccggaac cgcaggcacc ctcggagccc acccagatct gagctgtccc 780
tgatctcttc tagaggggaa gagcctattc cgtcccctac tccaagagca gagccattct 840
ccgcaaacgg cagcccccaa actgagctcc ctcccacaga actgtctgtc cacaccccat 900
ccccccaagc agaacctcta agccctgaaa ctgctcagac agaggtggcc cccagaacca 960
ggcctgcccc cctacggcat caccccagag cccaggcctc tggcacagag ccccccctcac 1020
ccacgcactc cttaggagaa ggtggcttct tccgtgcatc ccctcagcca cgaaggccaa 1080
gttcccaggg ttgggccagt ccccaggtag cagggagacg ccctgatcct tttccttcgg 1140
tccctcgggg ccgaggccag cagggccaag ggccttgggg aacgggggg actcctcacg 1200
ggccccgcct ggagcctgac cctcagcacc cgggcgcctg gctgcccctg ctgagcaacg 1260
gccccсatgc cagctccctc tggagcctct ttgctcccag tagccctatt ccaagatgtt 1320
ctggggagag tgaacagcta agagcctgca gccaagcgcc ctgcccccct gagcagccag 1380
acccccgggc cctgcagtgc gcagccttta actcccagga attcatgggc cagctgtatc 1440
agtgggagcc cttcactgaa gtccagggct cccagcgctg tgaactgaac tgccggcccc 1500
gtggcttccg cttctatgtc cgtcacactg aaaaggtcca ggatgggacc ctgtgtcagc 1560
ctggagcccc tgacatctgt gtggctggac gctgtctgag ccccggctgt gatgggatcc 1620
ttggctctgg caggcgtcct gatggctgtg gagtctgtgg gggtgatgat tctacctgtc 1680
gccttgtttc ggggaacctc actgaccgag gggggcccct gggctatcag aagatcttgt 1740
ggattccagc gggagccttg cggctccaga ttgcccagct ccggcctagc tccaactacc 1800
tggcacttcg tggccctggg ggccggtcca tcatcaatgg gaactgggct gtggatcccc 1860
ctgggtccta cagggccggc gggaccgtct ttcgatataa ccgtcctccc agggaggagg 1920
gcaaagggga gagtctgtcg gctgaaggcc ccaccaccca gcctgtggat gtctatatga 1980
tctttcagga ggaaaaccca ggcgtttttt atcagtatgt catctcttca cctcctccaa 2040
tccttgagaa ccccaccccca gagcccсctg tcccccagct tcagccggag attctgaggg 2100
tggagccccc acttgctccg gcaccccgcc cagcccggac cccaggcacc ctccagcgtc 2160
aggtgcggat cccccagatg cccgccccgc cccatcccag gacacccctg gggtctccag 2220
ctgcgtactg gaaacgagtg ggacactctg catgctcagc gtcctgcggg aaaggtgtct 2280
ggcgccccat tttcctctgc atctcccgtg agtcgggaga ggaactggat gaacgcagct 2340
gtgccgcggg tgccaggccc ccagcctccc ctgaaccctg ccacggcacc ccatgccccc 2400
catactggga ggctggcgag tggacatcct gcagccgctc ctgtggcccc ggcacccagc 2460
accgccagct gcagtgccgg caggaatttg gggggggtgg ctcctcggtg cccccggagc 2520
gctgtggaca tctcccccgg cccaacatca cccagtcttg ccagctgcgc ctctgtggcc 2580
attgggaagt tggctctcct tggagccagt gctccgtgcg gtgcggccgg ggccagagaa 2640
gccggcaggt tcgctgtgtt gggaacaacg gtgatgaagt gagcgagcag gagtgtgcgt 2700
caggcccccc acagcccccc agcagagagg cctgtgacat ggggccctgt actactgcct 2760
ggttccacag cgactggagc tccaagtgct cagccgagtg tgggacggga atccagcggc 2820
gctctgtggt ctgccttggg agtggggcag ccactcgggc caggccaggg ggaagcagga 2880
gcaggaactg ggcagagctg tccaacagga agccggcccc ctgacatgcg cgcctgcagc 2940
ctggggccct gtgagagaac ttggcgctgg tacacagggc cctggggtga gtgctcctcc 3000
gaatgtggct ctggcacaca gcgtagagac atcatctgtg tatccaaact ggggacggag 3060
ttcaacgtga cttctccgag caactgttct cacctcccca ggccccctgc cctgcagccc 3120
tgtcaagggc aggcctgcca ggaccgatgg ttttccacgc cctggagccc atgttctcgc 3180
tcctgccaag gggaacgca gacacgggag gtccagtgcc tgagcaccaa ccagaccctc 3240
agcacccgat gccctcctca actgcggccc tccaggaagc gcccctgtaa cagccaaccc 3300
tgcagccagc gccctgatga tcaatgcaag gacagctctc cacattgccc cctggtggta 3360
caggcccggc tctgcgtcta cccctactac acagccacct gttgccgctc ttgcgcacat 3420
gtcctggagc ggtctcccca ggatccctcc tgaaaggggt ccggggcacc ttcacggttt 3480

tctgtgccac catcggtcac ccattgatcg gcccactctg aaccccctgg ctctccagcc 3540
tgtcccagtc tcagcaggga tgtcctccag gtgacagagg gtggcaaggt gactgacaca 3600
aagtgacttt cagggctgtg gtcaggccca tgtggtggtg tgatgggtgt gtgcacatat 3660
gcctcaggtg tgcttttggg actgcatgga tatgtgtgtg ctcaaacgtg tatcactttt 3720
caaaaagagg ttacacagac tgagaaggac aagacctgtt tccttgagac tttcctaggt 3780
ggaaaggaaa gcaagtctgc agttccttgc taatctgagc tacttagagt gtggtctccc 3840
caccaactcc agttttgtgc cctaagcctc atttctcatg ttcagacctc acatcttcta 3900
agccgcccctg tgtctctgac cccttctcat ttgcctagta tctctgcccc tgcctcccta 3960
attagctagg gctggggtca gccactgcca atcctgcctt actcaggaag gcaggaggaa 4020
agagactgcc tctccagagc aaggcccagc tgggcagagg gtgaaaaaga gaaatgtgag 4080
catccgctcc cccaccaccc cgcccagccc ctagccccac tccctgcctc ctgaaatggt 4140
tcccacccag aactaattta ttttttatta aagatggtca tgacaaatga aaaaaaaaaa 4200
aaaaaaataa aaaaacaaaa aaaaaaaata 4230

<210> 34
<211> 3699
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3473847CB1

<400> 34
cgcagtgtgc tggcaaagct tgactttccc agcaggccta tgtcataggt actgtggtct 60
ctacaataca gcagaggtat ctgaggctcc gagaggttga gtgacttgct catggctgca 120
caaccagtaa atattggagc tggaattcag gtccacggtt tcctggctcc aaagcccatg 180
atttttttccc tcaatttatt ctgactgggg catgggggag ggggtggcct ttgggcaggg 240
ccaccaggag cgaccaggcc cgtagagagc tgggtgcagg tacagaggaa aacctgttgt 300
cgagtgtggc ccgtagttcc cattttgcc tgaatggcac atttgaaagt gttatataac 360
catgtgaata ataatagttg gcctatatga gttttttaat ttgcttttttg gtccgcattt 420
ggtaacttct ttatcatcta ctatactctg ttgtgtctct tttgttgtaa tttgtaagta 480
ggggtgagat aaagtacacc tagggtttgc tgggtttctt ccatgtcatc atgttcctcc 540
ttgcatgggg ccaggatccg tggaggttgc ctggcaccta cgtggtggtg ctgaaggagg 600
agacccacct ctcgcagtca gagcgcactg cccgccgcct gcaggcccag gctgcccgcc 660
ggggatacct caccaagatc ctgcatgtct tccatggcct tcttcctggc ttcctggtga 720
agatgagtgg cgacctgctg gagctggcct tgaagttgcc ccatgtcgac tacatcgagg 780
aggactcctc tgtctttgcc cagagcatcc cgtggaacct ggagcggatt accccctccac 840
ggtaccgggc ggatgaatac cagccccccg acggaggcag cctggtggag gtgtatctcc 900
tagacaccag catacagagt gaccaccggg aaatcgaggg cagggtcatg gtcaccgact 960
tcgagaatgt gcccgaggag gacgggaccc gcttccacag acaggccagc aagtgtgaca 1020
gtcatggcac ccacctggca ggggtggtca gcggccggga tgccggcgtg gccaagggtg 1080
ccagcatgcg cagcctgcgc gtgctcaact gccaagggaa gggcacggtt agcggcaccc 1140
tcataggcct ggagtttatt cggaaaagcc agctggtcca gcctgtgggg ccactggtgg 1200
tgctgctgcc cctggcgggt gggtacagcc gcgtcctcaa cgccgcctgc cagcgcctgg 1260
cgagggctgg ggtcgtgctg gtcaccgctg ccggcaactt ccgggacgat gcctgcctct 1320
actccccagc ctcagctccc gaggtcatca cagttggggc caccaatgcc caggaccagc 1380
cggtgaccct ggggactttg ggaccaact ttggccgctg tgtggacctc tttgccccag 1440
gggaggacat cattggtgcc tccagcgact gcagcacctg ctttgtgtca cagagtggga 1500
catcacaggc tgctgcccac gtggctggca ttgcagccat gatgctgtct gccgagccgg 1560
agctcaccct ggccgagttg aggcagagac tgatccactt ctctgccaaa gatgtcatca 1620
atgaggcctg gttccctgag gaccagcggg tactgacccc caacctggtg gccgccctgc 1680
cccccagcac ccatggggca ggttggcagc tgttttgcag gactgtgtgg tcagcacact 1740
cggggcctac acggatggcc acagccatcg cccgctgcgc cccagatgag gagctgctga 1800
gctgctccag tttctccagg agtgggaagc ggcggggcga gcgcatggag gcccaagggg 1860
gcaagctggt ctgccgggcc cacaacgctt ttgggggtga gggtgtctac gccattgcca 1920
ggtgctgcct gctaccccag gccaactgca gcgtccacac agctccacca gctgaggcca 1980
gcatggggac ccgtgtccac tgccaccaac agggccacgt cctcacaggc tgcagctccc 2040
actgggaggt ggaggacctt ggcacccaca agccgcctgt gctgaggcca cgaggtcagc 2100
ccaaccagtg cgtgggccac agggaggcca gcatccacgc ttcctgctgc catgccccag 2160
gtctggaatg caaagtcaag gagcatggaa tcccggcccc tcaggagcag gtgaccgtgg 2220
cctgcgagga gggctggacc ctgactggct gcagtgccct ccctgggacc tcccacgtcc 2280
tgggggccta cgccgtagac aacacgtgtg tagtcaggag ccgggacgtc agcactacag 2340
gcagcaccag cgaagaggcc gtgacagccg ttgccatctg ctgccggagc cggcacctgg 2400
cgcaggcctc ccaggagctc cagtgacagc cccatcccag gatgggtgtc tggggagggt 2460
caagggctgg ggctgagctt taaaatggtt ccgacttgtc cctctctcag ccctccatgg 2520
cctggcacga ggggatgggg atgcttccgc ctttccgggg ctgctggcct ggcccttgag 2580

tggggcagcc tccttgcctg gaactcactc actctgggtg cctcctcccc aggtggaggt 2640
gccaggaagc tccctccctc actgtggggc atttcaccat tcaaacaggt cgagctgtgc 2700
tcgggtgctg ccagctgctc ccaatgtgcc gatgtccgtg ggcagaatga cttttattga 2760
gctcttgttc cgtgccaggc attcaatcct caggtctcca ccaaggaggc aggattcttc 2820
ccatggatag gggagggggc ggtaggggct gcagggacaa acatcgttgg ggggtgagtg 2880
tgaaaggtgc tgatggccct catctccagc taactgtgga gaagcccctg ggggctccct 2940
gattaatgga ggcttagctt tctggatggc atctagccag aggctggaga caggtgtgcc 3000
cctggtggtc acaggctgtg ccttggtttc ctgagccacc tttactctgc tctatgccag 3060
gctgtgctag caacacccaa aggtggcctg cggggagcca tcacctagga ctgactcggc 3120
agtgtgcagt ggtgcatgca ctgtctcagc caacccgctc cactaccccgg caggtacac 3180
attcgcaccc ctacttcaca gaggaagaaa cctggaacca gagggggcgt gcctgccaag 3240
ctcacacagc aggaactgag ccagaaacgc agattgggct ggctctgaag ccaagcctct 3300
tcttacttca cccggctggg ctcctcattt ttacgggtaa cagtgaggct gggaagggga 3360
acacagacca ggaagctcgg tgagtgatgg cagaacgatg cctgcaggca tggaactttt 3420
tccgttatca cccaggcctg attcactggc ctggcggaga tgcttctaag gcatggtcgg 3480
gggagagggc caacaactgt ccctccttga gcaccagccc cacccaagca agcagacatt 3540
tatcttttgg gtctgtcctc tctgttgcct ttttacagcc aacttttcta gacctgtttt 3600
gcttttgtaa cttgaagata tttattctgg gttttgtagc atttttatta atatggtgac 3660
tttttaaaat aaaaacaaac aaacgttgtc ctaaaaaaa 3699

<210> 35
<211> 2410
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3750004CB1

<400> 35
cctcagcagt ggccccttcc ctccacgggc tgccccggag ctcagtccca cccccctccgc 60
cgatgaggcc atgaggcacc gaacggacct gggccagaac ctcctgctct tcctgtgggc 120
cctgctgaac tgtggtttgg gggtcagtgc tcagggtccg ggcgagtgga ccccgtgggt 180
gtcctggacc cgctgctcca gctcctgcgg gcgtggcgtc tctgtgcgca gccggcgctg 240
cctccggctt cctggggaag aaccgtgctg ggagactcc catgagtacc gcctctgcca 300
gttgccagac tgccccccag gggctgtgcc cttccgagac ctacagtgtg ccctgtacaa 360
tggccgccct gtcctgggca cccagaagac ctaccagtgg gtgcccttcc atggggcgcc 420
caaccagtgc gacctcaact gcctggctga ggggcacgcc ttctaccaca gcttcggccg 480
cgtcctggac ggcaccgcct gcagcccggg tgcccagggg gtctgcgtgg ctggccgctg 540
ccttagcgcc ggctgtgatg ggttgttggg ctcgggtgcc ctcgaggacc gctgtggccg 600
ctgcggaggc gccaacgact cgtgcctttt cgtgcagcgc gtgtttcgtg acgccggtgc 660
cttcgctggg tactggaacg tgaccctgat ccccgagggc gccagacaca tccgcgtgga 720
acacaggagc cgcaaccacc tgggtatcct aggatcactg atggggggcg atgggcgcta 780
cgtgcttaat gggcactggg tggtcagccc accagggacc tacgaggcgg ccggcacgca 840
tgtggtctac acccgagaca cagggcccca ggagacattg caagcagccg ggcccacctc 900
ccatgacctg ctcctacagg tcctcctgca ggagcccaac cctggcatcg agtttgagtt 960
ctggctccct cgggagcgct acagcccctt ccaggctcgt gtgcaggccc tgggctggcc 1020
cctgaggcag cctcagcccc ggggggtgga gcctcagccc cccgcagccc ctgctgtcac 1080
ccctgcacag accccaacgc tggccccaga cccctgccca ccctgccctg acacccgcgg 1140
ccgcgcccac cgactactcc actattgcgg cagtgacttt gtgttccagg cccgagtgct 1200
gggccaccac caccaggccc aggagacccg ctatgaggtg cgcatccagc tcgtctacaa 1260
gaaccgctcg ccactgcggg cacgcgagta cgtgtgggcg ccaggccact gcccctgccc 1320
gatgctggca ccccaccggg actacctgat ggctgtccag cgtcttgtca gccccgacgg 1380
cacacaggac cagctgctgc tgccccacgc cggctacgcc cggccctgga gccctgcgga 1440
ggacagccgc atacgcctga ctgcccggcg ctgtcctggc tgagcccctg caggagcccc 1500
ggccacacac agcaagaaag atacatctga ccagcctcaa cgtcaacgta tttcccctct 1560
caccctggct tccaggcagc tctgaaatac gtcccacctg tgcagctatg tgactccctc 1620
ccacacacgc ttaagacacc tctgcatgca gtcaaagcca ctgtcacaag ccggcaggca 1680
ctggtgagga ggcactaagg agactctgac ttttatttcg cctctctcct tggctgccag 1740
gaagctcata gctatttata ctcagaaagt ttaacgctgc tttctttctc tttgcgcgcg 1800
tcacacttgc ttggagacac tgtcatgaac gagcatgaca ccctgctgcc ctgggtaccc 1860
agaagatcat ctgtttactt cccagacact gtgctgtctc tgctctctgc tactcacaca 1920
cacccctcatg tgtgaagggc agagacactg tcacaaacag gcatgcccct tagaagacat 1980
gcctaaccag gcactgtaac gtaccaacgt accaatttcc ccttttcccc tggctaccag 2040
gaaactcgga gacaatcttt tcagcctcag catttctggc tggatttcca cccatcaaca 2100
cgtgcttgct cctccttttt ttttttctg aggtagacct tgctctgtca cctaggctga 2160
agtgcggtgg tgcaatcatg gctcactgca gcctcaaatt cctgggctca agcgatcctc 2220

ccactcagct ccacagcagt ggaactcacg tgtgatcaca tgccggctaa tttaaatttt 2280
gtagagatgg gcttgtacgt gccaaatgtc tcactatggc tcaacatctc tgctgggtcc 2340
aagactgaat aggatgacat gatggtgtac cccttatcct tatttcagct ttaaaaattc 2400
taaaaaaaaa 2410

<210> 36
<211> 549
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4904126CB1

<400> 36
gggaggagag aaaagccatg gccgacaagg tcctgaagga gaagagaaag cagtttatcc 60
gttcagtggg cgaaggtaca ataaatggct tactgggtga attattggag acaagggtgc 120
tgagccagga agagatagag atagtaaaat gtgaaaatgc tacagttatg gataaggccc 180
gagctttgct tgactctgtt attcggaaag ggctccagc atgccaaatt tgcatcacat 240
acatttgtga agaagacagt cacctggcag ggacgctggg actctcagca ggtccaacat 300
ctggaaatca ccttactaca caagattctc aaatagtact tccttcctag gtaatgctgt 360
ttttaaagaa agagcattct ttgaaccgtg gcttcccgtg acattaatgt tgtaggatga 420
accacagtta aaggggctat gaagaattcc catagagtga tcatacaatt ttctttttgt 480
aatctattct gcttttgtag caactgtcaa aacagcttca ctatctatgt ctacattaaa 540
atttggaat 549

<210> 37
<211> 2755
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71268415CB1

<400> 37
ttgctaggag ggtggagttc atccacttat gatataaatg tctcttttta ttttttgcag 60
gcaacttttt gctccttcct acacagaaac ccattatact tcaagtggta accctcaaac 120
caccacacgg aaattggagg atcactgctt ttaccacggc acggtgaggg agacagaact 180
gtccagcgtc acgctcagca cttgccgagg aattagagga ctgattacgg tgagcagcaa 240
cctcagctac gtcatcgagc ccctccctga cagcaagggc caacacctta tttacagatc 300
tgaacatctc aagccgcccc cgggaaactg tgggttcgag cactccaagc ccaccaccag 360
ggactgggct cttcagttta cacaacagac caagaagcga cctcgcagga tgaaaaggga 420
agatttaaac tccatgaagt atgtggagct ttacctcgtg gctgattatt tagagtttca 480
gaagaatcga cgagaccagg acgccaccaa acacaagctc atagagatcg ccaactatgt 540
tgataagttt taccgatcct tgaacatccg gattgctctc gtgggcttgg aagtgtggac 600
ccacgggaac atgtgtgaag tttcagagaa tccatattct accctctggt cctttctcag 660
ttggaggcgc aagctgcttg cccagaagta ccatgacaac gcccaattaa tcacgggcat 720
gtccttccac ggcaccacca tcggcctggc ccccctcatg gccatgtgct ctgtgtacca 780
gtctggagga gtcaacatgg accactccga gaatgccatt ggcgtggctg ccaccatggc 840
ccacgagatg ggccacaact ttggcatgac ccatgattct gcagattgct gctcggccag 900
tgcggctgat ggtgggtgca tcatggcagc tgccactggg caccccttc ccaaagtgtt 960
caatggatgc aacaggaggg agctggacag gtatctgcag tcaggtggtg gaatgtgtct 1020
ctccaacatg ccagacacca ggatgttgta tggaggccgg aggtgtggga acgggtatct 1080
ggaagatggg gaagagtgtg actgtggaga agaagaggaa tgtaacaacc cctgctgcaa 1140
tgcctctaat tgtaccctga ggccgggggc ggagtgtgct cacggctcct gctgccacca 1200
gtgtaagctg ttggctcctg ggaccctgtg ccgcgagcag gccaggcagt gtgacctccc 1260
ggagttctgt acgggcaagt ctccccactg ccctaccaac ttctaccaga tggatggtac 1320
cccctgtgag ggcggccagg cctactgcta caacggcatg tgcctcacct accaggagca 1380
gtgccagcag ctgtggggac ccggagcccg acctgcccct gacctctgct tcgagaaggt 1440
gaatgtggca ggagacacct ttggaaactg tggaaaggac atgaatggtg aacacaggaa 1500
gtgcaacatg agagatgcga agtgtgggaa gatccagtgt cagagctctg aggcccggcc 1560
cctggagtcc aacgcggtgc ccattgacac cactatcatc atgaatggga ggcagatcca 1620
gtgccggggc acccacgtct accgaggtcc tgaggaggag ggtgacatgc tggacccagg 1680
gctggtgatg actggaacca agtgtggcta caaccatatt tgctttgagg ggcagtgcag 1740
gaacacctcc ttctttgaaa ctgaaggctg tgggaagaag tgcaatggcc atggggtctg 1800
taacaacaac cagaactgcc actgcctgcc gggctgggcc ccgccccttct gcaacacacc 1860

gggccacggg ggcagtatcg acagtgggcc tatgcccccct gagagtgtgg gtcctgtggt 1920
agctggagtg ttggtggcca tcttggtgct ggcggtcctc atgctgatgt actactgctg 1980
cagacagaac aacaaactag gccaactcaa gccctcagct ctcccttcca agctgaggca 2040
acagttcagt tgtcccttca gggtttctca gaacagcggg actggtcatg ccaacccaac 2100
tttcaagctg cagacgcccc agggcaagcg aaaggtgatc aacactccgg aaatcctgcg 2160
gaagccctcc cagcctcctc cccggccccc tccagattat ctgcgtggtg ggtccccacc 2220
tgcaccactg ccagctcacc tgagcagggc tgctaggaac tccccagggc ccgggtctca 2280
aatagagagg acggagtcgt ccaggaggcc tcctccaagc cggccaattc cccccgcacc 2340
aaattgcatc gtttcccagg acttctccag gcctcggccg ccccagaagg cactcccggc 2400
aaacccagtg ccaggccgca ggagcctccc caggccagga ggtgcatccc cactgcggcc 2460
ccctggtgct ggccctcagc agtcccggcc tctggcagca cttgccccaa aggtgagtcc 2520
acgggaagcc ctcaaggtga aagctggtac cagagggctc caggggggca ggtgtagagt 2580
tgagaaaaca aagcaattca tgcttcttgt ggtctggact gaacttccag aacaaaagcc 2640
aagggcaaaa cattcatgtt tcttggtgcc cgcttgactg tggagttttg gcttcatgtg 2700
aaaggtgatt cttagaatcc tgagctgtgg tggcttcagt cctgcccctg cacct 2755

<210> 38
<211> 2553
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473301CB1

<400> 38
atggacaaag aaaacagcga tgtttcagcc gcacctgctg acctgaaaat atccaatatc 60
tcagtccaag tggtcagtgc ccaaaagaag ctgccagtga gacgaccacc gttgccaggg 120
agacgactac cattgccagg aagacgacca ccacaaagac ccattggcaa agccaaaccc 180
aagaagcaat ccaagaaaaa agttcccttt tggaatgtac aaaataaaat cattctcttc 240
acagtatttt tattcatcct agcagtcata gcctggacac ttctgtggct gtatatcagt 300
aagacagaaa gcaaagatgc tttttacttt gctgggatgt ttcgcatcac caacatcgag 360
tttcttcccg aataccgaca aaaggagtcc agggaatttc tttcagtgtc acggactgtg 420
cagcaagtga taaacctggt ttatacaaca tctgccttct ccaaatttta tgagcagtct 480
gttgttgcag atgtcagcag caacaacaaa ggcggcctcc ttgtccactt ttggattgtt 540
tttgtcatgc cacgtgccaa aggccacatc ttctgtgaag actgtgttgc cgccatcttg 600
aaggactcca tccagacaag catcataaac cggacctctg tggggagctt gcagggactg 660
gctgtggaca tggactctgt ggtactaaat ggtgattgtt ggtcattcct aaaaaaaaag 720
aaaagaaagg aaaatggtgc tgtctccaca gacaaaggct gctctcagta cttctatgca 780
gagcatctgt ctctccacta cccgctggag atttctgcag cctcagggag gctgatgtgt 840
cacttcaagc tggtggccat agtgggctac ctgattcgtc tctcaatcaa gtccatccaa 900
atcgaagccg acaactgtgt cactgactcc ctgaccattt acgactccct tttgcccatc 960
cggagcagca tcttgtacag aatttgtgaa cccacaagaa cattaatgtc atttgtttct 1020
acaaataatc tcatgttggt gacatttaag tctcctcata tacggaggct ctcaggaatc 1080
cgggcatatt ttgaggtcat tccagaacaa aagtgtgaaa acacagtgtt ggtcaaagac 1140
atcactggct ttgaagggaa aatttcaagc ccatattacc cgagctacta tcctccaaaa 1200
tgcaagtgta cctggaaatt tcagacttct ctatcaactc ttggcatagc actgaaattc 1260
tataactatt caataaccaa gaagagtatg aaaggctgtg agcatggatg gtgggaaatt 1320
tatgagcaca tgtactgtgg ctcctacatg gatcatcaga caatttttcg agtgcccagc 1380
cctctggttc acattcagct ccagtgcagt tcaaggcttt caggcaagcc acttttggca 1440
gaatatggca gttacaacat cagtcaaccc tgccctgtgg gatcttttag atgctcctcc 1500
ggtttatgtg tccctcaggc ccagcgtggt gatggagtaa atgactgctt tgatgaaagt 1560
gatgaactgt tttgcgtgag ccctcaacct gcctgcaata ccagctcctt caggcagcat 1620
ggccctctca tctgtgatgg cttcagggac tgtgagaatg gccgggatga gcaaaactgc 1680
actcaaagta ttccatgcaa caacagaact tttaagtgtg gcaatgatat ttgctttagg 1740
aaacaaaatg caaaatgtga tgggacagtg gattgtccag atggaagtga tgaagaaggc 1800
tgcacctgca gcaggagttc ctccgccctt caccgcatca tcggaggcac agacaccctg 1860
gaggggggtt ggccgtggca ggtcagcctc cactttgttg gatctgccta ctgtggtgcc 1920
tcagtcatct ccagggagtg gcttctttct gcagcccact gttttcatgg aaacaggctg 1980
tcagatccca caccatggac tgcacacctc gggatgtatg ttcaggggaa tgccaagttt 2040
gtctccccgg tgagaagaat tgtggtccac gagtactata acagtcagac ttttgattat 2100
gatattgctt tgctacagct cagtattgcc tggcctgaga ccctgaaaca gctcattcag 2160
ccaatatgca ttcctcccac tggtcagaga gttcgcagtg gggagaagtg ctgggtaact 2220
ggctgggggc gaagacacga agcagataat aaaggctccc tcgttctgca gcaagcggag 2280
gtagagctca ttgatcaaac gctctgtgtt tccacctacg ggatcatcac ttctcggatg 2340
ctctgtgcag gcataatgtc aggcaagaga gatgcctgca aaggagattc gggtggacct 2400
ttatcttgtc gaagaaaaag tgatggaaaa tggattttga ctggcattgt tagctgggga 2460

catggatgtg gacgaccaaa ctttcctggt gtttacacaa gggtgtcaaa ctttgttccc 2520
tggattcata aatatgtccc ttctcttttg taa 2553

<210> 39
<211> 1041
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473308CB1

<400> 39
atgttcagcg gcaacacagg aaaaacccat attatcaatg ctcaaaaacc tggccacctc 60
aggcttagcc agttattcgt gagcagagag gtgtgtcatc tacatggcag tcatggcctg 120
gatgggtctg gaactgtggc aagaatcctt ccaggaaaca gccggtctcc ctctctgctc 180
tcagaaggca agtttcctta tcacctgtct gctctcagaa ggcaagtttc cttatcacct 240
gtgaatcaca aacccacaga gtggccaaac atactgatgc aagaccatag gaaggggaaa 300
gctgcagttg gtgtctcctt tgatgatgat gacaagattg ttgggggcta caactgtgag 360
gagaattctg tcccctacca ggtgtccctg aattctggct accacttctg tgttggctcc 420
ctcaacaggg aatactgcat ccaggtgaga ctgggagagc acaacatcga agtcctagag 480
gggaatgaac agttcatcta tgcggtcaag atcatccgcc accccaaata caacagctgg 540
actctggaca atgacatcct gctgatcaag ctctccacac ctgccatcat caatgcccat 600
gtgtccacca tctctctgcc caccacccct ccagctgctg gcactgagtg cctcatctct 660
ggctggggca acactctgag ttctggcgcc gactacccag acgagctgca gtgcctggat 720
gctcctgtgc tgagccaggc tgagtatgaa gcctcctacc ctggaaagat taccaacaac 780
gtgttttgtg tgggtttcct tgagggaggc aaggattcct gccagattat tcctatcaaa 840
gtgcagcagc tggttacctc aagccaagag acagacataa ggatccctat ggccttgcag 900
acagctgctt ccacctccta cctgggcccc ttagactctt tacacaggaa agtgagtcac 960
cccactgaga agcgttgcca gcagaaacag ggcatgaaaa tcacagataa ccatgggatt 1020
acttccaagt ggtcagtata a 1041

<210> 40
<211> 1707
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7478021CB1

<400> 40
atgctcgccg cctccatctt ccgtccgaca ctgctgctct gctggctggc tgctccctgg 60
cccacccagc ccgagagtct cttccacagc cgggaccgct cggacctgga gccgtcccca 120
ctgcgccagg ccaagcccat tgccgacctc cacgctgctc agcggttcct gtccagatac 180
ggctggtcag ggggtgtggc ggcctggggg cccagtcccg aggggccgcc ggagaccccc 240
aagggcgccg ccctggccga ggcggtgcgc aggttccagc gggcgaacgc gctgccggcc 300
agcggggagc tggacgcggc caccctagcg gccatgaacc ggccgcgctg cggggtcccg 360
gacatgcgcc caccgccccc ctccgccccg ccttcgcccc cgggcccgcc ccccagagcc 420
cgctccaggc gctccccgcg ggcgccgctg tccttgtccc ggcggggttg gcagccccgg 480
ggctaccccg acggcggagc tgcccaggcc ttctccaaga ggacgctgag ctggcggctg 540
ctgggcgagg ccctgagcag ccaactgtcc gtggccgacc agcggcgcat agaggcgctg 600
gccttcagga tgtggagcga ggtgacgccg ctggacttcc gcgaggacct ggccgccccc 660
ggggccgcgg tcgacatcaa gctgggcttt gggagacggc acctgggctg tccgcgggcc 720
ttcgatggga gcgggcagga gtttgcacac gcctggcgcc taggtgacat tcactttgac 780
gacgacgagc acttcacacc tcccaccagt gacacgggca tcagccttct caaggtggcc 840
gtccatgaaa ttggccatgt cctgggcttg cctcacacct acaggacggg atccataatg 900
caaccaaatt acattcccca ggagcctgcc tttgagttgg actggtcaga caggaaagca 960
attcaaaagc tgtatggttc ctgtgaggga tcatttgata ctgcgtttga ctggattcgc 1020
aaagagagaa accaatatgg agaggtgatg gtgagattta gcacatattt cttccgtaac 1080
agctggtact ggctttatga aaatcgaaac aataggacac gctatgggga ccctatccaa 1140
atcctcactg gctggcctgg aatcccaaca cacaacatag atgcctttgt tcacatctgg 1200
acatggaaaa gagatgaacg ttattttttt caaggaaatc aatactggag atatgacagt 1260
gacaaggatc aggccctcac agaagatgaa caaggaaaaa gctatcccaa attgatttca 1320
gaaggatttc ctggcatccc aagtcccccta gacacggcgt tttatgaccg aagacagaag 1380
ttaatttact tcttcaagga gtcccttgta tttgcatttg atgtcaacag aaatcgagta 1440
cttaattctt atccaaagag gattactgaa gtttttccag cagtaatacc acaaaatcat 1500

cctttcagaa atatagattc cgcttattac tcctatgcat acaactccat tttctttttc 1560
aaaggcaatg catactggaa ggtagttaat gacaaggaca aacaacagaa ttcctggctt 1620
cctgctaatg gcttatttcc aaaaaagttt atttcagaga agtggtttga tgtttgtgac 1680
gtccatatct ccacactgaa catgtaa 1707

<210> 41
<211> 1262
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4333459CB1

<400> 41
aaaagatctt tgcgaaacac tacattcaga aacatcagat ggacatgctt gattcaccac 60
gtcttggtta atgaataaac ttgttttaaa ttggcttatt gctggtctct caaggcttcc 120
tatttttgtt tgctttagtc tctctaaaat ttcagggaaa aactatgagt ctcaaaatgc 180
ttataagcag gaacaagctg atttactac taggaatagt ctttttgaa cgaggtaaat 240
ctgcaactct ttcgctcccc aaagctccca gttgtgggca gagtctggtt aaggtacagc 300
cttggaatta ttttaacatt ttcagtcgca ttcttggagg aagccaagtg gagaaggtt 360
cctatccctg gcaggtatct ctgaaacaaa ggcagaagca tatttgtgga ggaagcatcg 420
tctcaccaca gtgggtgatc acggcggctc actgcattgc aaacagaaac attgtgtcta 480
ctttgaatgt tactgctgga gagtatgact taagccagac agacccagga gagcaaactc 540
tcactattga aactgtcatc atacatccac atttctccac caagaaacca atggactatg 600
atattgccct tttgaagatg gctggagcct tccaatttgg ccactttgtg gggcccatat 660
gtcttccaga gctgcgggag caatttgagg ctggttttat ttgtacaact gcaggctggg 720
gccgcttaac tgaaggtggc gtcctctcac aagtcttgca ggaagtgaat ctgcctattt 780
tgacctggga agagtgtgtg gcagctctgt taacactaaa gaggcccatc agtgggaaga 840
cctttctttg cacaggtttt cctgatggag ggagagacgc atgtcaggga gattcaggag 900
gttcactcat gtgccggaat aagaaagggg cctggactct ggctggtgtg acttcctggg 960
gtttgggctg tggtcgaggc tggagaaaca atgtgaggaa aagtgatcaa ggatcccctg 1020
ggatcttcac agacattagt aaagtgcttt cctggatcca cgaacacatc caaactggta 1080
actaagccat cacacaaggt taagaagctg ccattctgct agggccagag acagcatcag 1140
cagagtcctg gcaaatcaga gcacctgaac caacaggctc tacctctgtt ctcagtgtag 1200
cacacaagga ttgtgaggtt taccaagtct aaataaaaca agagttaaat atggtaaaaa 1260
aa 1262

<210> 42
<211> 3067
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6817347CB1

<400> 42
gcactgtgaa cgttggttgc atccaaatct gaattttgtc tgggaccagg gtcagggacc 60
agaatacacc agagctgagg gccagcccta cctgagaacc atcaacaaac ttaccccaca 120
tcccattata cctcctcact ccctgcagcc tgtcagcttc cccaatctcc cacactcact 180
gtcacctggg gctctggtgc accagatgac actacttgct ccctggtaca caggccccat 240
gatccccatg gatgttaatg agcccagctc cgtgaccacg gctcctaccc tcagctctag 300
cctgcagcat atctcctcat tcctggccac tggtaagaaa ctttccctcc attttggtca 360
tccacgtgag tgtgaagtca ccaggattga tgacaaaaat agaagaggat tggaagacag 420
tgagccaggt gccaaactct tcaataatga tggagtctgt tgttgcctgc aaaaacgggg 480
gccagtgaac attacatcag tgtgtgtgag tcccaggacc ttacaaatat cagtttttgt 540
gttatcagag aaatacgagg gtattgttaa atttgaatcg gatgaattac cttttggtgt 600
aattggttct aatattggtg atgcacattt tcaagaattc agggctggaa tctcctggaa 660
gcctgtggta gatcctgatg accccattcc tcagttccct gattgctgca gcagcagcag 720
cagcaggatt ccttcagtga gtgtgctagt tgcagttcct ctggttgcag gccacaaagg 780
gcaggcattt attgaaagga tgctggggtg cttcaaggaa ttgaagcaag agctgactca 840
ggaagggccg ggcgggggac accccaggtc tgcgtggccc ccgcgccgcc acgcccagtg 900
gccgccccgag ccctgcgagc agggggagga gccgccgcca gtggaggcgg aggaggtaga 960
ggaggcggag acggcggaga aggcggagag gaaggtggag gcggaggcga aggtggaggg 1020
gaaggcggag gcggcgggga aggcggaggc ggcggggaag gtggacgcca ccgagaaggt 1080
ggagacggcg gggaaggtgg acgccgctgg gaaggtggag acggcggagg gtccgggccg 1140

```
ccgggctgag ctcaagctgg agcccgaacc cgagccggtc cgggaggcgg agcaggagcc 1200
gaagcaggag ctggaggatg agaacccagc gcggagcggc ggtggcggca acagcgacga 1260
ggttcctccc cccacccttc cctccgatcc accgcggccc cccgatccct ctccgcgtcg 1320
cagtcgtgcg ccgcgccgcc gaccccggcc ccggccccag acccggctcc gtaccccgcc 1380
gcagcctagg ccccggcccc cgccccggcc ccggccccgg cgcggccctg gggcggatg 1440
cctggatgtg gattttgccg tggggccacc aggctgttct cacgtgaaca gctttaaggt 1500
gggagagaac tggaggcagg aactgcgggt tatctaccag tgcttcgtgt ggtgtggaac 1560
cccagagacc aggaaaagca aggcaaagtc ctgcatctgc catgtgtgtg gcacccatct 1620
gaacagactc cactcttgcc tttcctgtgt cttctttggc tgcttcacgg agaaacacat 1680
tcacgagcac gcagagacga aacaacacaa cttagcagta gacctgtatt acggaggtat 1740
atactgcttt atgtgtaagg actatgtata tgacaaagac attgagcaaa ttgccaaaga 1800
agagcaagga gaagctttga aattacaagc ctccacctca acagaggttt ctcaccagca 1860
gtgttcagtg ccaggccttg gtgagaaatt cccaacctgg gaaacaacca aaccagaatt 1920
agaactgctg gggcacaacc cgaggagaag aagaatcacc tccagcttta cgatcggttt 1980
aagaggactc atcaatcttg gcaacacgtg ctttatgaac tgcattgtcc aggccctcac 2040
ccacacgccg atactgagag atttctttct ctctgacagg caccgatgtg agatgccgag 2100
tcccgagttg tgtctggtct gtgagatgtc gtcgctgttt cgggagttgt attctggaaa 2160
cccgtctcct catgtgccct ataagttact gcacctggtg tggatacatg cccgccattt 2220
agcagggtac aggcaacagg atgcccacga gttcctcatt gcagcgttag atgtcctgca 2280
caggcactgc aaaggtgatg atgtcgggaa ggcggccaac aatcccaacc actgtaactg 2340
catcatagac caaatcttca caggtggcct gcagtctgat gtcacctgtc aagcctgcca 2400
tggcgtctcc accacgatag acccatgctg ggacattagt ttggacttgc ctggctcttg 2460
cacctccttc tggcccatga gcccagggag ggagagcagt gtgaacgggg aaagccacat 2520
accaggaatc accaccctca cggactgctt gcggaggttt acgaggccag agcacttagg 2580
aagcagtgcc aaaatcaaat gtggtagttg ccaaagctac caggaatcta ccaaacagct 2640
cacaatgaat aaattacctg tcgttgcctg ttttcatttc aaacggtttg aacattcagc 2700
gaaacagagg cgcaagatca ctacatacat ttcctttcct ctggagctgg atatgacgcc 2760
gtttatggcc tcaagtaaag agagcagaat gaatggacaa ttgcagctgc caaccaatag 2820
tggaaacaac gaaaataagt attccttgtt tgctgtggtt aatcaccaag gaaccttgga 2880
gagtggccac tataccagct tcatccggca ccacaaggac cagtggttca agtgtgatga 2940
tgccgtcatc actaaggcca gtattaagga cgtactggac agtgaagggt atttactgtt 3000
ctatcacaaa caggtgctag aacatgagtc agaaaaagtg aaagaaatga acacacaagc 3060
ctactga                                                           3067
```